Université de Montréal

ESSAYS IN EMPIRICAL FINANCE

par
Magnim Farouh

Département de sciences économiques
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures et postdoctorales
en vue de l'obtention du grade de PhilosophiæDoctor (Ph.D.)
en Sciences économiques

Août, 2020

Université de Montréal
Faculté des études supérieures et postdoctorales

Cette thèse intitulée:

ESSAYS IN EMPIRICAL FINANCE

présenté par:
Magnim Farouh

a été évaluée par un jury composé des personnes suivantes:

Guillaume Sublet	Président-rapporteur
René Garcia	Directeur de recherche
Vasia Panousi	Co-directrice de recherche
Marine Carrasco	Membre du jury
Ruslan Goyenko	Examinateur externe
....	Représentant du doyen de la FAS

Thèse accepté le ....

Résumé

Cette thèse comporte trois chapitres dans lesquels j'étudie les coûts de transaction des actions, les anomalies en finance et les activités du système bancaire parallèle.

Dans le premier chapitre (*co-écrit avec René Garcia*), une nouvelle façon d'estimer les coûts de transaction des actions est proposée. Les coûts de transaction ont diminué au fil du temps, mais ils peuvent augmenter considérablement lorsque la liquidité de financement se raréfie, lorsque les craintes des investisseurs augmentent ou lorsqu'il y a d'autres frictions qui empêchent l'arbitrage. Nous estimons dans ce chapitre les écarts entre les cours acheteur et vendeur des actions de milliers d'entreprises à une fréquence journalière et présentons ces mouvements importants pour plusieurs de ces épisodes au cours des 30 dernières années. Le coût de transaction des trois quarts des actions est fortement impacté par la liquidité de financement et augmente en moyenne de 24 %. Alors que les actions des petites entreprises et celles des entreprises à forte volatilité ont des coûts de transaction plus élevés, l'augmentation relative des coûts de transaction en temps de crise est plus prononcée pour les actions des grandes entreprises et celles des entreprises à faible volatilité. L'écart entre les coûts de transaction respectifs de ces groupes de qualité élevée et qualité faible augmente également lorsque les conditions financières se détériorent, ce qui prouve le phénomène de fuite vers la qualité. Nous avons construit des portefeuilles basés sur des anomalies et avons estimé leurs "alphas" ajustés pour les coûts de rééquilibrage sur la base de nos estimations des coûts de transaction pour montrer que toutes les stratégies sont soit non rentables soit perdent de l'argent, à l'exception de deux anomalies: le "prix de l'action" et la "dynamique du secteur industriel".

Dans le deuxième chapitre, j'étudie comment la popularité des anomalies dans les revues scientifiques spécialisées en finance peut influer sur le rendement des stratégies basées sur ces anomalies. J'utilise le ton du résumé de la publication dans laquelle une anomalie est discutée et le facteur d'impact de la revue dans laquelle cette publication a paru pour prévoir le rendement des stratégies basées sur ces anomalies sur la période après publication. La principale conclusion est la suivante: lorsqu'une anomalie est discutée dans une publication dont le résumé a un ton positif, et qui apparaît dans une revue avec un facteur d'impact supérieur à 3 (Journal of Finance, Journal of Financial Economics, Review of Financial Studies), cette anomalie est plus susceptible d'attirer les investisseurs qui vont baser leurs stratégies sur cette anomalie et corriger ainsi la mauvaise évaluation des actions.

Le troisième chapitre (*co-écrit avec Vasia Panousi*) propose une mesure de l'activité bancaire parallèle des entreprises opérant dans le secteur financier aux États-Unis. À cette fin, nous utilisons l'analyse de données textuelles en extrayant des informations des rapports annuels et trimestriels des entreprises. On constate que l'activité bancaire parallèle était plus élevée pour les "Institutions de dépôt", les "Institutions qui ne prennent pas de dépôt" et le secteur "Immobilier" avant 2008. Mais après 2008, l'activité bancaire parallèle a considérablement baissé pour toutes les firmes opérant dans le secteur financier sauf les "Institutions non dépositaires". Notre indice du système bancaire parallèle satisfait certains faits économiques concernant le système bancaire parallèle, en particulier le fait que les politiques monétaires restrictives contribuent à l'expansion du système bancaire parallèle. Nous montrons également avec notre indice que, lorsque l'activité bancaire parallèle des 100 plus grandes banques augmente, les taux de délinquance sur les prêts accordés par ces banques augmentent également. L'inverse est observé avec l'indice bancaire traditionnel: une augmentation de l'activité bancaire traditionnelle des 100 plus grandes banques diminue le taux de délinquance.

Mots clés: Coûts de transaction, Anomalie, Liquidité de financement, Échantillonnage de Gibbs, Facteur de Bayes, Analyse de données textuelles, Facteur d'impact d'un journal, Activité bancaire parallèle

Abstract

This thesis has three chapters in which I study transaction costs, anomalies and shadow banking activities.

In the first chapter (*co-authored with René Garcia*) a novel way of estimating transaction costs is proposed. Transaction costs have declined over time but they can increase considerably when funding liquidity becomes scarce, investors' fears spike or other frictions limit arbitrage. We estimate bid-ask spreads of thousands of firms at a daily frequency and put forward these large movements for several of these episodes in the last 30 years. The transaction cost of three-quarters of the firms is significantly impacted by funding liquidity and increases on average by 24%. While small firms and high volatility firms have larger transaction costs, the relative increase in transaction costs in crisis times is more pronounced in large firms and low-volatility firms. The gap between the respective transaction costs of these high- and low-quality groups also increases when financial conditions deteriorate, which provides evidence of flight to quality. We build anomaly-based long-short portfolios and estimate their alphas adjusted for rebalancing costs based on our security-level transaction cost estimates to show that all strategies are either unprofitable or lose money, except for price per share and industry momentum.

In the second chapter I study how the popularity of anomalies in peer-reviewed finance journals can influence the returns on these anomalies. I use the tone of the abstract of the publication in which an anomaly is discussed and the impact factor of the journal in which this publication appears to forecast the post-publication return of strategies based on the anomaly. The main finding is the following: when an anomaly is discussed in a positive tone publication that appears in a journal with an impact factor higher than 3 (Journal of Finance, Journal of Financial Economics, Review of Financial Studies), this anomaly is more likely to attract investors that are going to arbitrage away the mispricing.

The third chapter (*co-authored with Vasia Panousi*) proposes a measure of the shadow banking activity of firms operating in the financial industry in the United States. For this purpose we use textual data analysis by extracting information from annual and quarterly reports of firms. We find that the shadow banking activity was higher for the "Depository Institutions", "Non depository Institutions" and the "Real estate" before 2008. But after 2008, the shadow banking activity dropped considerably for all the financial companies except for the "Non depository Institutions". Our shadow banking index satisfies some economic facts about the shadow banking, especially the fact that contractionary monetary policies contribute to expand shadow banking. We also show with our index that, when the shadow banking activity of the 100 biggest banks increases, the delinquency rates on the loans that these banks give also increases. The opposite is observed with the traditional banking index: an increase of the traditional banking activity of the 100 biggest banks decreases the delinquency rate.

Keywords: Transaction cost, Anomaly, Funding liquidity, Gibbs sampling, Bayes factor, Textual analysis, Journal impact factor, Shadow banking, Term frequency - Inverse document frequency.

Table des matières

Table of contents

Liste des tableaux

List of Tables

Liste des Figures

List of Figures

à mes parents, Abra et Yolome.

Remerciements

Acknowledgements

J'aimerais tout d'abord remercier mon directeur de recherche, René Garcia, pour la confiance qu'il a placée en moi en acceptant de m'encadrer mais surtout, pour sa disponibilité, son suivi continu, ses nombreux conseils et sa gentillesse. Merci René, d'avoir été mon mentor pendant toutes ces années.

Je remercie par ailleurs Vasia Panousi, pour tout le soutien qu'elle m'a apporté dans les dernières années de mes études doctorales. Elle est le professeur avec qui j'ai appris les techniques d'analyse textuelle qui m'ont beaucoup servi dans ma thèse.

Toute ma gratitude à William Mccausland pour sa disponibilité, à Benoît Perron pour ses remarques et conseils, à tous les professeurs du département d'économie de l'Université de Montréal, ainsi qu'à l'ensemble du personnel administratif dudit département.

Mes sincères remerciements à tous les doctorants en économie de l'Université de Montréal, et plus particulièrement à mes camarades de promotion Fatim, Idriss, Lionel, Lucienne, Marlène, N'Golo, et Samuel ainsi qu'à Félicien, Kokouvi, et Souleymane. Merci pour votre présence et vos encouragements. Je remercie par ailleurs mon grand frère Germain, mon père et ma mère pour leurs soutien et conseils dans les moments de doute.

Je tiens enfin à remercier le Département de sciences économiques de l'Université de Montréal, la Faculté des arts et des sciences, le Centre Inter-universitaire de Recherche en Economie Quantitative (CIREQ) et la Faculté des études supérieures et postdoctorales, pour leur soutien financier et logistique.

Avant-propos

Foreword

Expliquer l'évolution des rendements des actifs financiers a toujours été une préoccupation pour les chercheurs en finance. Sharpe [1964] et Lintner [1965] ont posé les bases de ce qui est aujourd'hui un champ de recherche florissant en évaluation des actifs financiers: l'explication des rendements à partir d'un nombre réduit de facteurs observés. Le CAPM a été le premier modèle de ce genre, et a été suivi par une multitude de modèles à facteurs dont les plus populaires sont en particulier le modèle de Fama-French, le CAPM intertemporel, le CAPM conditionnel, et le modèle de Carhart, etc..

Une anomalie est définie comme toute variable qui procure une prédictibilité inconsistante avec le modèle d'évaluation d'actifs financiers considéré (CAPM, modèles Fama-French ou autre). Ces dernières années, les chercheurs en finance et comptabilité ont publié dans les revues de finance et comptabilité plus de 150 anomalies. Les anomalies ont donné aux investisseurs l'opportunité de pouvoir construire des stratégies pour pouvoir réaliser des profits. Ces stratégies consistent à construire des portefeuilles dynamiques d'actions exposés à l'anomalie. Le dynamisme de ces portefeuilles requiert des ajustements fréquents qui font que les investisseurs font face à des coûts de transaction sur le marché financier, surtout en période de crise de liquidité financière.

Pour pouvoir profiter d'une anomalie, l'investisseur doit d'abord être au courant qu'une telle anomalie existe et pour cela il doit s'informer en se renseignant sur les avancées dans le domaine de la recherche académique en finance. Une fois qu'il est au courant de l'anomalie, l'investisseur doit créer le portefeuille dynamique basé sur cette anomalie. Avoir une bonne mesure des coûts de transaction sur le marché financier est ainsi vital pour un investisseur qui aimerait baser ses stratégies sur les anomalies, étant donné que maintenir un portefeuille constamment exposé à l'anomalie nécessite des ajustements impliquant ainsi des coûts de transaction. Une bonne mesure du coût de transaction d'une action devrait inclure les éléments relatifs à cette action, mais aussi les éléments relatifs à l'environnement global du marché financier et économique.

L'environnement économique et financier dépend fortement des activités bancaires. Le secteur bancaire procure aux agents économiques (dont ceux qui investissent sur le marché financier) du financement pour pouvoir mener leurs activités. Dans le secteur bancaire, nous distinguons deux types de banques : les banques traditionnelles qui reçoivent des dépôts et accordent des prêts aux ménages et aux entreprises, sous la supervision des régulateurs et des banques centrales; les banques parallèles, par contre sont des intermédiaires financiers qui facilitent la création de crédit dans l'économie via la titrisation des actifs, sans accepter de dépôts et sans faire l'objet d'une surveillance réglementaire. Quelques exemples de banques parallèles incluent les fonds spéculatifs, les compagnies d'assurance et les sociétés dérivées. Vu l'importance de leurs activités durant la crise financire de 2008, les banques parallèles semblent être celles qui comportent le plus de risques. Avoir un bon suivi des activités des banques parallèles est donc très important pour les décideurs publics.

Les trois articles de cette thèse s'inscrivent dans une logique de développement de nouveaux outils permettant de:

- mesurer les coûts de transaction des actions et l'effet de ces coûts sur les profits de stratégies basées sur les anomalies;
- voir comment les investisseurs sélectionnent les anomalies en fonction des journaux dans lesquels ces anomalies sont publiées;
- avoir une nouvelle mesure des activités du secteur bancaire parallèle.

Dans le premier chapitre de cette thèse, une estimation des coûts de transactions des actifs financiers prenant en compte les frictions qui peuvent constituer un frein pour l'arbitrage, est proposée. Le deuxième chapitre montre comment la popularité des anomalies publiées dans les revues scientifiques spécialisées en finance et la qualité de ces revues peuvent influer sur le rendement des stratégies basées sur ces anomalies. Le troisième chapitre, enfin, propose une mesure des activités bancaires parallèles basée sur les rapports annuels des firmes qui opèrent dans le secteur financier aux États-Unis.

Chapter 1

Financial Risks, Transaction costs and Performance of Anomalies

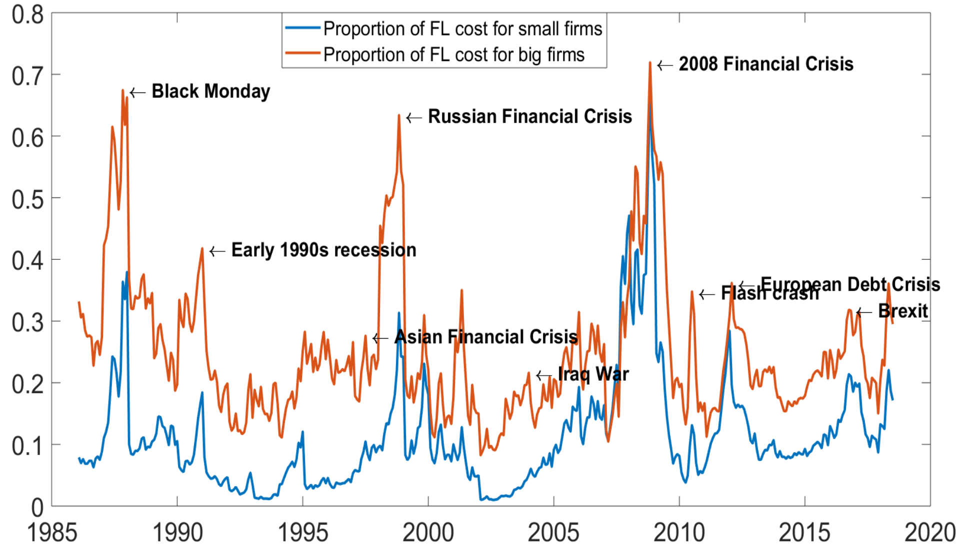
1.1 Introduction

Algorithmic trading is everywhere present in financial markets. While the trading rules are set by fund managers, their execution is fully automatized[1]. A positive effect of this automation is the reduction of transaction costs. Machines help reduce the part of transaction costs that is related to a firm's specific information since news are instantly reflected in its security price. However, transaction costs of firms may be also affected by aggregate market conditions such as limited funding liquidity, heightened investors' fears or other frictions that limit arbitrage.

The main focus of this paper is to provide estimates of transaction costs that include the cost of these market frictions. We incorporate measures of financial risks such as funding liquidity or tail risk in the current estimation procedures of bid-ask spreads. A second objective is to measure the impact of these time-varying, market-based transaction costs on the returns of long-short strategies that arbitrageurs are pursuing by building portfolios sorted on firm characteristics. We assess what remains of the alpha of dynamic strategies based on so-called anomalies once we incorporate the additional rebalancing costs due to aggregate frictions.

[1] According to a recent article in The Economist, funds run by computers that follow rules set by humans account for 35% of America's stock market, 60% of institutional equity assets and 60% of trading activity. According to Deutsche Bank, 90% of equity-futures trades and 80% of cash-equity trades are executed by algorithms without any human input.

Figure 1.1: **Proportion of funding liquidity t-cost for small firms and big firms**



In our model, the transaction cost is written as an affine function of the financial risk measure: t-cost $= c_0 + c_1 \times$ financial-risk. For the estimation we use the Gibbs Sampling method of Hasbrouck [2004] and Hasbrouck [2009]. Once we have the estimates of c_0 and c_1 for a given stock and a given year, we can compute the round-trip transaction cost for this stock, for a given day t, by t-cost$_t = 2 \times (c_0 + c_1 \times$ financial-risk$_t)$. To test if the financial risk variable is statistically relevant for estimating the transaction cost we compute the Bayes factor between the Hasbrouck [2009] model (hereafter H-model) and the extended model including the financial risk variable (hereafter FLH-model for funding liquidity, TRH-model for tail risk and VIX-H model for the VIX).

Our empirical results support funding liquidity as an important factor in the estimation of transaction costs. To illustrate the role played by funding conditions especially in crisis periods, Figure 1.1 plots the time series of the proportion of the transaction costs due to funding liquidity for large and small firms. For each big financial market event between 1986 and 2018, the proportion jumps to about 60-70 % for large firms. Overall, large firms are relatively more impacted by funding conditions than small firms since their transaction costs are small in normal times. However, in the 2008 financial crisis that raised considerably liquidity risk, the proportion for small and large firms are about the same. Over the period from January 1986 to June 2018, we find that there is more evidence for the FLH-Model against the H-Model for 73% of firm-years. Estimated transaction costs are in average 24% higher for the FLH-Model compared to the H-Model. The tail risk and VIX variables are also important in the estimation of transaction costs. We find with the Bayes factor that, there is more evidence for the TRH-Model and the VIX-H model against the H-Model respectively for 86% of firm-years and 74% of firm-years. With respect to the H-model, the estimated transaction costs are 19% and 95% higher for the TRH- and VIX-H models respectively.

We also investigate whether the estimated transaction costs reflect flight to quality. The quality of a particular stock is positively related to the size of the firm (Lang and Lundholm [1993]) and negatively related to the stock's volatility (Brunnermeier and Pedersen [2009]). We will say that there is evidence for flight to quality if the differential in transaction costs between high and low quality stocks increases when the financial risk increases. We find that the differentials in transaction costs between small and large firms and high- and low-volatility firms increase when the financial risk increases.

The estimated transaction costs are also used to assess the after-trading-cost performance of long-short anomaly-based portfolios. The latter are constructed for a set of anomalies considered one at a time. Each month, stocks are ranked based on the value of the anomaly. Stocks are then grouped into deciles. The long-short portfolio is then obtained by going long on the stocks in the highest decile and short on the stocks in the lowest decile or inversely, depending on the anomaly[2]. Given the way the portfolios are built, each month or each year depending on the trading frequency, the stocks included in a given decile are not necessarily the same as in the previous month or year. Therefore, to stay exposed to the anomaly, the portfolios need to be rebalanced and transaction costs are incurred. We find that a proper accounting of the adjusted transaction costs for financial risks eliminates the profits of a large number of anomaly-based long-short portfolios.

For robustness purposes, we also consider the estimation of transaction costs with the model of Lesmond et al. [1999]. This model requires only the time series of daily security returns to endogenously estimate the effective transaction costs for any firm, exchange, or time period. The feature of the data that allows for the estimation of transaction costs is the incidence of zero returns. We introduce funding liquidity in this model and perform a likelihood ratio test with a model without frictions. The model with funding liquidity is

[2]Let us cite two examples. For momentum, the portfolio is obtained by going long on the stocks in the highest decile and short on those in the lowest decile. For size, it is the reverse. The portfolio is long on stocks in the lowest decile and short on stocks in the highest decile.

preferred to the model without funding liquidity for a third of the firms.

The main objectives of the paper are motivated by two recent contributions to the literature on transaction costs. Weller [2019] uses equity bid-ask spreads at high-frequency to infer a measure of tail risk, while Patton and Weller [2019] introduce market liquidity and funding liquidity variables to determine whether trading strategies based on some characteristics are implementable in practice. They better capture the price impact of trading strategies in large portfolios compared to Novy-Marx and Velikov [2016].

Our paper differs from this recent literature both in its methodology and scope. To measure the bid-ask spread of individual securities, we extend the model of Hasbrouck [2009] by adding financial risk variables to the market return factor. Our main application is based on the TED spread (short-term LIBOR minus short Treasury rate) that is used to measure funding liquidity cost. We also consider for robustness purposes the measure of tail risk proposed by Weller [2019] and the VIX to capture liquidity frictions at the aggregate level.

Adding these variables to the measure of transaction costs is supported theoretically. Brunnermeier and Pedersen [2009] propose a theoretical model that shows that transaction costs depend on funding liquidity. Since Weller [2019] proposes a measure of tail risk based on the cross-section of bid-ask spreads, such a measure can help recover effective transaction costs. This is explained by the fact that liquidity providers, in moments of tight funding constraints or extreme events require a high compensation leading to high transaction costs. The VIX could also affect transaction costs because higher volatility tightens funding constraints of market makers and thereby reduces their liquidity-provision capacity (Gromb and Vayanos [2002], Brunnermeier and Pedersen [2009],Nagel [2012]).

The rest of the paper is structured as follows. After reviewing the literature on the measurement of transaction costs, we describe in Section 1.2 the estimation methodology of transaction costs based on the model of Hasbrouck [2009] and the extensions made to include measures of financial risks. We also explain how the new model including financial risk is compared to the basic Hasbrouck model. Section 1.3 describes the data used for the estimation. Section 1.4 presents the results of the transaction costs estimation for the various models. Robustness checks are reported in Section 1.5. In Section 1.6 the performances of anomaly-based strategies are computed after taking in account the transaction costs augmented by financial frictions. Section 1.7 concludes.

Related Literature

The most direct way to measure transaction costs is to take the bid-ask spread plus commissions. However, according to Ng et al. [2008], the bid-ask spread underestimates the real transaction costs because it does not take in account relevant elements such as price impact or opportunity costs. According to Roll [1984], commissions depend on a number of hard-to-quantify factors (such as the transaction size, the amount of business done by the investor, and the time of day or year) given that they are negotiated. Grossman and Miller [1988] argue that, for a given trade, it is unlikely that the seller and the buyer arrive at the same time on the market and thus the spread cannot serve as the measure of the transaction cost. Another issue with the bid-ask spread is that it is not always available for all firms and time periods where security returns exist.

To overcome the issues associated with a direct measure of bid-ask spreads based on trades and quotes, Roll [1984] proposed a model to estimate the transaction costs by a so-called effective bid-ask spread. The suggested measure for the effective bid-ask spread is based on the fact that transaction costs induce negative serial dependance in successive observed market price changes. However, it is not always the case that this covariance is negative in the data. To overcome this issue, Hasbrouck [2004] proposes a Gibbs sampling estimate of Roll [1984] model that is based on daily closing prices. Hasbrouck [2009] extends the Hasbrouck [2004] model by including a market return factor in the estimation equation and shows that the estimated effective spreads have a 96.5% correlation with the ones estimated from actual trades from the trade and quote (TAQ) dataset. Goyenko et al. [2009] confirms that the effective bid-ask spread is a good proxy for the bid-ask spreads estimated with intra-daily trade-and-quote data.

The Bayesian procedure proposed by Hasbrouck [2009] necessitates long time series, leaving some firms without a transaction cost estimate. A solution is to rely on proxies.[3] Novy-Marx and Velikov [2016] use the fact that market capitalization and idiosyncratic volatility explain around 70% of the cross section of transaction costs to assign transaction costs to stocks for which the model proposed by Hasbrouck [2009] could not deliver an estimate.

Lesmond et al. [1999] propose a model of security returns that avoids the limitations of

[3] Karpoff and Walkling [1988] and Bhushan [1994] use price, trading volume, firm size, and the number of shares outstanding, variables assumed to be negatively related to transaction costs. Of course, proxy variables may capture effects that are not due to transaction costs and cannot be used to compute net returns of a portfolio.

the transaction cost proxies. The effect of transaction costs is modeled through the incidence of zero returns. If the value of the information signal is insufficient to exceed the costs of trading, then the marginal investor will not trade, causing a zero return. The estimates from this model are the marginal traders effective transaction costs. Lesmond et al. [2004] use this methodology to compute the after-transaction-cost returns of different momentum portfolios to prove that the profits from momentum strategies are illusory.

The implementation costs of financial market anomalies has also been studied recently by Patton and Weller [2019]. They estimate the transaction costs of mutual funds strategies by relying on Corwin and Schultz [2011]'s methodology to estimate bid-ask spreads based on daily high and low prices.

Our paper is also related to the literature on the link between market liquidity (as measured by the bid-ask spread) and financial risk measures such as funding liquidity (Gromb and Vayanos [2002], Brunnermeier and Pedersen [2009] and Kondor and Vayanos [2019]), the VIX (Nagel [2012]) or tail risk (Weller [2019]). Aragon and Strahan [2012] document empirically the relationship between funding liquidity and market liquidity by linking the market liquidity of stocks held by hedge funds exposed to Lehman Brothers to shocks to funding liquidity during the bankruptcy.

Our paper also relates to the large literature about the limits of arbitrage (Shleifer and Vishny [1997], Geanakoplos [2010], Gromb and Vayanos [2010]). Tight funding conditions increase transaction costs and therefore prevent arbitrageurs from taking advantage of mis-priced assets.

1.2 Methodology

To overcome the issues associated with a direct measure of bid-ask spreads based on trades and quotes, Roll [1984] proposed a model to estimate the transaction costs by a so-called effective bid-ask spread from daily security prices. In this section we describe the estimation procedures to arrive at a measure of transaction costs that fluctuates with a measure of financial risk. Since it is a Bayesian estimation procedure we provide all the steps of the Gibbs-sampling algorithm for the various parameters of the model.

1.2.1 Measuring the Effective Bid-ask Spread from Daily Prices

To incorporate funding liquidity risk, volatility risk or tail risk into the measure of the effective bid-ask spreads of firms, we extend the Bayesian procedure of Hasbrouck [2009]. We start by describing the model of Roll [1984] on which the procedure is based, then the Bayesian estimation and finally the incorporation of the financial risk factor in the procedure.

a. The model of Roll (1984)

Transaction prices are composed of a random-walk and a noise, wherein the random-walk is the "efficient price" of security and the noise is the bid-ask spread, as follows:

$$m_t = m_{t-1} + \varepsilon_t \tag{1.1}$$

$$b_t = m_t - c \tag{1.2}$$

$$a_t = m_t + c \tag{1.3}$$

where m_t is the efficient price', b_t the bid price and a_t the ask price, all expressed in logarithms, ε_t a random disturbance reflecting public information about the stock, and c is the half-spread, presumed to reflect the quote-setter's cost of market-making.

The model introduces a random indicator q_t to capture the direction of the trade. It takes the value one with probability 0.5 if the trade takes place at the ask, and minus one with probability 0.5 if it does at the bid.

If p_t is the observed transaction price, equations (1.2) and (1.3) can be summarized with the following equation: $p_t = m_t + c.q_t$. Therefore:

$$\Delta p_t = c\Delta q_t + \varepsilon_t, \tag{1.4}$$

which yields $Cov(\Delta p_t, \Delta p_{t+1}) = Cov(c.\Delta q_t + \varepsilon_t, c.\Delta q_{t+1} + \varepsilon_{t+1})$. In most implementations of the Roll model, it is assumed that the direction of the trade is independent of the efficient price movement i.e. q_t is independent of ε_t. With this assumption, we obtain $Cov(\Delta p_t, \Delta p_{t+1}) = c^2.Cov(\Delta q_t, \Delta q_{t+1})$. Given that q_t is equal to $+1$ or -1 with equal probabilities, $Cov(\Delta p_t, \Delta p_{t+1}) = -c^2$. Therefore, the half-spread is equal to $c = \sqrt{-Cov(\Delta p_t, \Delta p_{t+1})}$.

This way of estimating c is infeasible when we have positive auto-covariances between daily changes in stock prices. Roll [1984] finds that auto-covariance estimates based on 21 daily returns are positive for almost half the cases. Harris [1990] studies the statistical properties of the Roll bid-ask spread estimator and shows that positive auto-covariances are more likely for low values of the spread.

Hasbrouck [2009] argues that another problem arises when there is no trade on a particular day. When there is no trade on a particular day, CRSP reports the midpoint of the closing bid and ask. If these days are retained in the sample, the estimated cost will generally be biased downward, because the midpoint realizations do not include the cost. If these days are dropped from the sample, heteroscedasticity may arise since the efficient price innovations may span multiple days.

b. The Bayesian procedure of Hasbrouck (2004, 2009)

To overcome this issue, Hasbrouck [2004] proposes a Bayesian approach. In this approach, Hasbrouck [2004] makes two key assumptions: the spread is positive and $\varepsilon_t iid \sim N(0, \sigma_\varepsilon^2)$. The model parameter set is $\Theta = \{\sigma_\varepsilon^2, c\}$. Denote the prior parameter density as $\pi(\Theta)$. The posterior is given by $f(\Theta/p) = \frac{f(p/\Theta).\pi(\Theta)}{f(p)}$, where $p = \{p_1, p_2, ..., p_T\}$ denotes the vector of observed prices.

This posterior cannot be directly evaluated because the data likelihood function $f(\Theta/p)$ involves the unobserved $q = \{q_1, q_2, ..., q_T\}$. The problem is solved by considering $f(\Theta, q/p)$ and then by integrating out the q. Hasbrouck [2004] uses a Markov-Chain Monte-Carlo (MCMC) approach for this purpose.

Hasbrouck [2009] extends the model by including a market return factor in the Roll model:

$$\Delta m_t = \beta_m r_{mt} + \varepsilon_t. \tag{1.5}$$

Therefore, if we replace m_t by $p_t + c.q_t$, the observed price change is given by:

$$\Delta p_t = c\Delta q_t + \beta_m r_{mt} + \varepsilon_t. \tag{1.6}$$

The assumption $\varepsilon_t \ iid \sim N(0, \sigma_\varepsilon^2)$ is maintained, while the new parameter set is $\Theta = \{\sigma_\varepsilon^2, c, \beta\}$. The problem is solved by considering $f(\Theta, q/p.r_m)$ and then by integrating out the q using a MCMC approach, like in the original paper.

1.2.2 The Hasbouck Model with Financial Risk

In the model of Hasbrouck [2009], the transaction cost of a firm for a time period (be it a month, a quarter or a year) is estimated from its daily returns. It means that the cost will be constant for each time period. Our contribution is to link the transaction cost to a financial risk measure and make it time-varying at the daily level.

1.2.3 Estimation of the effective bid-ask spread with financial risk

We write the transaction cost as an affine function of the financial risk measure FR_t. We make the notation more precise than in the previous sections since we have to distinguish the time scales for the various coefficients and identify the firm since we will be forming anomaly portfolios in the second part of the paper.

$$m_t^i = m_{t-1}^i + \varepsilon_t^i \tag{1.7}$$

$$p_t^i = m_t^i + (c_{0,t_p}^i + c_{1,t_p}^i.FR_t)q_t^i, \tag{1.8}$$

where m_t^i is the underlying log efficient value, p_t^i is the log trade price, q_t^i is a random indicator for the direction of the trade that takes the value one (minus one) if the trade took place at the ask (bid), ε_t^i is a random disturbance reflecting public information about the stock, and $c_t^i = c_{0,t_p}^i + c_{1,t_p}^i.FL_t$ is the effective cost of trading. The subscript t corresponds to the daily frequency, while t_p denotes the time period over which we estimate the transaction cost (monthly or yearly). The coefficients c_{0,t_p} and c_{1,t_p} are two coefficients that are constant over each period p but vary from period to period. The effective cost of trading for firm i will be time varying at the daily level. The number of firms will be different each day and will be denoted by n_t.

By generalizing the previous equation to include a market return factor, as in Hasbrouck [2009], we obtain the following equation:

$$\Delta p_t^i = (c_{0,t_p}^i + c_{1,t_p}^i.FR_t)\Delta q_t^i + \beta_m^i r_{mt} + \varepsilon_t^i \tag{1.9}$$

This equation can be rewritten as:

$$\Delta p_t^i = (c_{0,t_p}^i.\Delta q_t^i + c_{1,t_p}^i.FR_t.\Delta q_t^i + \beta_m^i r_{mt} + \varepsilon_t^i \tag{1.10}$$

As Hasbrouck [2009], we need to follow a Bayesian approach to estimate this model since q_t^i, the random indicator for the direction of the trade, is unknown. We also assume that ε_t^i is iid $N(0, \sigma_{\varepsilon^i}^2)$. The parameters that will be estimated are c_0^i, c_1^i, β_m^i and $\sigma_{\varepsilon^i}^2$.

a. Simulating the Coefficients in a Linear Regression

The standard Bayesian normal regression model is $y = Xb + e$ where y is a column vector of n observations of the dependent variable, X is an $(n \times k)$ matrix of fixed regressors, b is a vector of coefficients, and the residuals are zero-mean multivariate normal $e \sim N(0, \Omega_e)$. Given Ω_e and a normal prior on b, $b \sim N(\mu_b, \Omega_b)$, the posterior is $b \sim N(\mu_b^*, \Omega_b^*)$, where $\mu_b^* = (X'\Omega_e^{-1}X + \Omega_b^{-1})^{-1}(X'\Omega_e^{-1}y + \Omega_b^{-1}\mu_b)$ and $\Omega_b^* = (X'\Omega_e^{-1}X + \Omega_b^{-1})^{-1}$.

In our framework, the linear regression we have is $\Delta p_t^i = c_0^i.\Delta q_t^i + c_1^i \cdot FR_t \cdot \Delta q_t^i + \beta_m^i r_{mt} + \varepsilon_t^i$. Non-negativity is imposed on c_0^i and c_1^i in order to keep the transaction cost $c_t = c_0^i + c_1 \cdot FR_t$ positive, since any of the financial risk measures considered is positive.

b. Simulating the Error Covariance Matrix

We also make the same assumption for $\Omega_e = \sigma^2 I$ than Hasbrouck [2009]. The prior distribution for σ^2 is an inverted gamma distribution: $\sigma^2 \sim IG(\alpha, \beta)$. The posterior distribution will also be an inverted gamma $\sigma^2 \sim IG(\alpha^*, \beta^*)$, where $\alpha^* = \alpha + \frac{n}{2}$ and $\beta^* = [\beta^{-1} + \sum \frac{e_i}{2}]^{-1}$.

c. Simulating the Trade Direction Indicators

The remaining step in the sampler involves drawing $q = q_1, ..., q_T$ when c_0, c_1, β_m, and σ^2 are known. The procedure is the same as the one used in Hasbrouck [2009]. The procedure is sequential. The first draw is $q_1/q_2, ..., q_T$, the second draw is $q_2/q_1, q_3, q_4, ..., q_T$, the third draw is $q_3/q_1, q_2, q_4, ..., q_T$, etc., where the "/" stands for the conditional draw.

d. Steps of the Sampling Procedure

For the sampler, we follow the steps and simulation parameter choices used in Hasbrouck [2009].

- Step 0 (initialization). Although the limiting behavior of the sampler is invariant to starting values, "reasonable" initial guesses may hasten convergence. The trade direction indicators q_t that do not correspond to midpoint reports are set to the sign of the most recent price change, with q_1 set (arbitrarily) to +1 and those corresponding to midpoint reports are set to 0; σ_ε^2 is initially set to 0.0004[4]. No initial values are required for c_0, c_1 and β_m, as they are drawn first.

- Step 1. Based on the most recently simulated values for σ_ε^2 and the set of q_t, compute the posterior for the regression coefficients (c_0, c_1 and β_m) and make a new draw.

- Step 2. Given c_0, c_1 and β_m, and the set of q_t, compute the implied ε_t, update the posterior for σ_ε^2, and make a new draw.

- Step 3. Given c_0, c_1, β_m and σ_ε^2 , make draws for $q_1, q_2, ..., q_T$. q_t that correspond to midpoint reports are not drawn and are equal to 0. Go to Step 1.

Each sampler is run for 1,000 sweeps.[5] Of the 1,000 draws for each parameter, the first 200 are discarded to burn in the sampler by removing the effect of starting values. The average of the remaining 800 draws (an estimate of the posterior mean) is used as a point estimate of the parameter.

e. A Bayes Factor to compare the Extended Model with Financial Risk to the Hasbrouck Model

The Bayes factor is the ratio of the marginal likelihoods of both models. Let $M_{1,i,y}$ and $M_{2,i,y}$ denote the marginal likelihoods of the Hasbouck model and the extended model, respectively, and D the data set. For a firm i and a year y, the ratio can be written as:

$$BF_{i,y} = \frac{P(M_{2,i,y}/D)}{P(M_{1,i,y}/D)} = \frac{P(M_{2,i,y}).P(D/M_{2,i,y})}{P(M_{1,i,y}).P(D/M_{1,i,y})} \tag{1.11}$$

[4]This roughly corresponds to a 30% annual idiosyncratic volatility

[5]We ran the estimation with 5,000 and 10,000 sweeps and obtained similar results.

Once again, the presence of the latent variable q_t complicates the computation of the marginal likelihoods. For this purpose, we use the reciprocal importance sampling of Gelfand and Dey [1994].

Let the prior density of Θ_k (assumed to be proper) be given by $\pi(\Theta_k/M_{k,i,y})$ and let $\Theta_k^{(m)} = \{\Theta_k^{(1)}, ..., \Theta_k^{(M)}\}$ be M draws from the posterior density $\pi(\Theta_k/D, M_{k,i,y})$ obtained using a Gibbs Sampler. Gelfand and Dey [1994] show that

$$\widehat{m}_{GD,k} = \left\{ \frac{1}{S} \sum_{s=1}^{S} \left(\frac{p(\Theta_k^{(s)})}{f(\Theta_k/D, M_{k,i,y}).\pi(\Theta_k/M_{k,i,y})} \right) \right\}^{-1}, \tag{1.12}$$

converges to $m(D/M_{k,i,y})$.

Therefore, we compute the Bayes factor with the following formula:

$$BF_{i,y} = \frac{\widehat{m}(D/M_{2,i,y})}{\widehat{m}(D/M_{1,i,y})}. \tag{1.13}$$

By referring to Jeffreys [1998], there is strong evidence for $M_{2,i,y}$ against $M_{1,i,y}$ if $BF_{i,y} > 10^{\frac{3}{2}}$.

1.3 Data Construction

As detailed in the previous section, we need daily returns of all stocks and of the market and a daily series of the financial risk measures, to estimate the transaction costs of all firms. We obtain the individual stock and market returns from the Center for Research in Security Prices (CRSP) database where each security has a unique identifier (PERMNO). The financial risk variables used in this paper are the TED spread, the VIX and the tail risk measure proposed by Weller [2019]. The TED spread and the VIX were downloaded from the Federal Reserve Economic Data (FRED) of the Federal Reserve Bank of St. Louis. The TED spread series (TEDRATE) spans the period from January 1986 to June 2018, while the VIX series (CBOE Volatility Index) runs from January 1990 to June 2018. The daily tail risk measure was obtained by aggregating the hourly tail risk measures in Weller [2019] from January 2008 to December 2014.[6] Therefore, the transaction costs corresponding to the three financial risk variables are estimated over the same respective samples.

[6]We are thankful to Brian Weller for providing us with the hourly series of tail risk.

To compute performance of long-short anomaly-based portfolios, we first construct anomalies following Novy-Marx and Velikov [2016] and Kozak et al. [2019] from two data sources: COMPUSTAT (North America - Fundamentals Quarterly) and CRSP. The list of anomalies and their description is provided in Appendix 1.7. To build portfolios for each anomaly say a, we start from all firms[7] for which the anomaly's value is available at each date t[8] and sort them according to this value. We separate the firms into deciles and compute the average return at a monthly frequency. If the value of the anomaly is available at a frequency lower than a month, say a year, the composition of each decile portfolio is kept the same for all the months in this year. The average return of a portfolio is computed from the monthly returns (from CRSP) by value-weighting them. The before-trading-cost performance of the portfolios is measured using the *alpha* from the Fama-French three-factor model.[9]

1.4 Transaction Costs with Funding Liquidity Risk

Brunnermeier and Pedersen [2009] propose a model where market liquidity and funding liquidity cause each other and are mutually reinforcing, potentially leading to liquidity spirals. When funding liquidity conditions are tight, traders are reluctant to take on capital intensive positions in high-margin securities, which lowers market liquidity. Similarly, when market liquidity is low, it becomes riskier to finance a trade and intermediaries ask for higher margins. While market liquidity is defined as the difference between the transaction price and the fundamental value (that is the transaction cost in Hasbrouck [2009] model) and is therefore measurable, funding liquidity is referred to as the shadow cost of capital and is latent. To measure funding liquidity at the daily frequency, we follow Frazzini and Pedersen [2014]) and use the TED spread, that is the difference between the three-month Treasury bill rate and the three-month LIBOR (London Interbank Offered Rate) in US dollars.[10] An increase in the TED spread signals that lenders believe default risk is increasing and funding conditions are getting tight.

[7] Our sample have about 260,000 firm-years with 27,000 firms traded on the NYSE, AMEX and NASDAQ stock exchanges.

[8] Date could be a year, a month, or a quarter, depending on the anomaly

[9] Data on Fama French 3 factors are obtained from the Data Library of Kenneth French website.

[10] Other interest-rate spreads have been used in the literature. Garleanu and Pedersen [2011] measure the shadow cost of capital bythe LIBORgeneral collateral (GC) repo interest-rate spread, while Park [2015] use the Libor-Overnight Index Swaps (OIS) spread. Several other measures are available at lower frequency (see Fontaine and Garcia [2012], Hu et al. [2013], and Golez et al. [2018]).

1.4.1 Average Transaction Costs

In Table 1.1, we report the average transaction costs for all stocks that we considered in our database. Over the period from 1986 to 2018, the average transaction cost is about 3% for the H-Model and 3.4% for the FLH-Model, while the difference between the medians is of the same magnitude. The standard deviations of the two models are close. The skewness and the kurtosis are very large for the two models.

Table 1.1 and Table 1.2 here.

The average of the round-trip transaction costs for the anomaly-based decile portfolios are reported in Table 1.2. For each portfolio we take the simple average of the estimated transaction costs of the firms included in the portfolio. We also report the statistic Pr which is the percentage of cases where the Bayes factor favors the model with funding liquidity.

For certain anomalies, we can see a large difference in the transaction costs between the extreme decile portfolios. For three anomalies that are related to the size of the firm (SIZE or market capitalization, PRICE and NOA or Net Operating Assets), there is a difference of 500 basis points between the portfolio of small firms (D1) and the portfolio of large firms (D10).[11] Adding funding liquidity increases only marginally this difference, but the statistic Pr indicates that there is more evidence for the FLH-model as size increases (from 0.52 for D1 to 0.90 for D10). It means that when funding liquidity conditions get tighter the relative impact on the liquidity of large securities is more pronounced than for small firms, which are more illiquid at all times.

These averages hide the nonlinear relationship between the effective bid-ask spread and the measure of funding liquidity. Figure 1.2 illustrates this relationship for small and big firms. We plot the average difference between the transaction costs of the two models ($Tcost_{FLH-Model} - Tcost_{H-Model}$) for big firms and for small firms against the level of the TED spread. Each dot of the scatter plot represents a month. Whether it is for big firms or for small firms, the difference in transaction increases with the TED spread but more so for large values of the spread, that is when funding conditions worsen. The nonlinear effect is

[11] Hirshleifer et al. [2004] document that high normalized net operating assets is associated with a rising trend in earnings that is not subsequently sustained. High Net Operating Assets stocks are more attractive thus have a higher market liquidity than low Net Operating Assets stocks. The inverse relationship between the transaction cost and the price are consistent with Bhushan [1994] who uses share price to proxy for the inverse of transaction costs.

more pronounced for smaller firms. Note that for low values of the spread some differences are close to zero or negative. This corresponds to the good times where funding liquidity is not significantly related to the effective bid-ask spread and it is more often the case for small firms.

Figure 1.2 here.

For the realized volatility portfolios, there is a 8% difference in the effective bid-ask spread between the low-volatility portfolio (D1) and the high-volatility portfolio (D10). We note the same monotonic pattern in the Pr statistic as for the size of the firm, it decreases as the volatility increases (from 0.80 to 0.52). In other words the relative impact of tight funding conditions is more pronounced for the low volatility portfolios. Nevertheless, the absolute difference between the transaction costs of the two models is higher for the high volatility portfolios.

Another sizable difference between the extreme decile portfolios is noted for momentum anomalies. The losers portfolio (D1) exhibit a much larger effective bid-ask spreads than the winners (D10), with a difference of 320 basis points for MOM11 and 260 basis points for MOM6. The Bayes factor statistic support the funding liquidity model with proportions from 0.63 to 0.81. The long-term reversal (LTREV), momentum reversal (MOMREV) and return on assets (ROAA) anomalies have somewhat important differences in transaction costs between the extreme portfolios in the order of 200 to 300 basis points, with a strong support for the funding liquidity model.

Overall, for the other anomalies, there are smaller differences between the effective bid-ask spreads of the extreme decile portfolios, and the model with funding liquidity is always supported by the Bayes factor with proportions higher than 60%.

1.4.2 The Dynamics of Transaction Costs

The averages we just discussed hide strong trends in historical transaction costs and marked spikes around crisis periods where trading frictions occur. In this section we will examine these dynamics for three anomalies that exhibited the largest difference in averages between the two extreme decile portfolios: size, realized volatility, and momentum.

a. Transaction costs and firm size

In Figures 1.3 and 1.4, we plot the transaction costs estimated with the Hasbrouck model (H-Model) at an annual frequency and the extended model with funding liquidity (FLH-Model), which shows the monthly movements associated with the TED spread, for a large firm (COCA COLA CO, with a market capitalization of 200 billions US dollars in 2018) and a small firm (ROCKY MOUNTAIN CHOCOLATE FACTORY, with a market capitalization of 22 millions US dollars in 2018). These two individual securities capture both the historical trends in transaction costs and the large fluctuations associated with trading frictions and captured by the TED spread. For both the large firm, COCA-COLA, and the small firm, ROCKY MOUNTAIN, we note a declining trend over time in transaction costs, from around 200 to 50 basis points for COCA-COLA and from about 750 to 100 basis points for ROCKY MOUNTAIN. This fact is known, but what is less documented is the huge spikes that occur when tight funding conditions impair trading. In the market crash of 1987 and the financial crisis of 2008, the transaction costs spiked at values between 500 and 650 basis points for COCA-COLA and between 1000 and 1500 basis points for ROCKY MOUNTAIN. The annual average estimates of the Hasbrouck [2009] model obscure these large fluctuations in trading costs.

Figure 1.3 and Figure 1.4 here.

Figure 1.5 and Figure 1.6 feature the evolution over the 1986 to 2018 period of the transaction costs estimated monthly with both models for size portfolios (D1 for small firms and D10 for large firms). Figure 1.5 confirms the downward time trend for the average of small firms from 10% in 1986 to about 2% in 2018. Two large spikes appear. The 2008 financial crisis is of course one of the two but the largest one occurred in April 1992. In fact it is a culmination since it follows the recessionary period of 1990-1991 that brought the transaction cost to a level of 18% at the end of 1992 from a level of around 10% at the beginning of 1990. Monthly estimates from both models follow closely each other, with the liquidity estimate above the fixed cost of Hasbrouck [2009].

Figure 1.5 and Figure 1.6 here.

In Figure 1.6 for large firms, a downward is also apparent between the beginning and the end of the sample, but during the decade 1990-2000 we observe a steady increase from

1992 to the beginning of 2000 for the liquidity model estimate of the transaction cost. It corresponds to an increase of 0.7 in the TED spread from the end of 1992 to the middle of year 2000. The peak at around 350 basis points for the liquidity model and 300 basis points for the H-Model occurred in the first months of the year 2000, coincident with the large jump in valuation during the tech bubble.

b. Transaction costs and volatility

Brunnermeier and Pedersen [2009] link market liquidity (that is the transaction price minus the fundamental value, in other words our measured transaction cost) to fundamental volatility. The link is connected with margin constraints and is stronger when funding conditions are tight. High-volatility securities are more affected by intermediaries' wealth shocks.

We measure these relations between the transaction costs of the individual stocks and their realized volatility. In Table 1.2, we report the average round-trip transaction costs for the realized-volatility decile portfolios. We observe that the high-volatility portfolio has a much higher transaction cost (around 9%) than the low-volatility portfolio (about 1%)[12]. Funding liquidity adds another 40 points in average for the trading cost of high-volatility stocks. The Bayes factor selects the funding liquidity model in a higher proportion (around 80%) for lower-volatility deciles than for higher volatility stocks (about 65%).

Figure 1.7 and Figure 1.8 show the average transaction costs for high-volatility stocks and low-volatility stocks over time. We note the same declining trend in the transaction cost of the high-volatility stocks from about 10% in 1986 to 4% in 2018. However, as already noted for the small firms, we observe a large increase from the end of 1987 to 1994. The crash plus the recession of 1990-1991 and the jump in the funds rate in 1994 made high-volatility firms more expensive to trade (to more than 20%). The second peak appears of course during the 2008 financial crisis. For the low-volatility firms, the relative difference between the two models is more pronounced than for the high-volatility firms and varies between 20 and 50 basis points except for the crash of 1987 and the financial crisis of 2008, with spreads of 90 and 150 basis points respectively.

Figure 1.7 and Figure 1.8 here.

[12]High-volatility stocks for a given year are the stocks that fall in the highest decile when we rank all stocks according to their realized volatility while low-volatility stocks are the stocks that fall in the lowest decile

c. Transaction Costs and Flight to Quality

Stocks of large firms and of low-volatility firms can be characterized as high-quality firms. In Brunnermeier and Pedersen [2009], flight to quality occurs when the market liquidity differential between high- and low-quality securities is larger bigger when speculator funding is tight. To rephrase this assertion, we can say that the flight to quality is the fact that the transaction cost differential between high- and low-quality securities (stocks of big and small firms or low-volatility and high-volatility stocks) is larger bigger when funding conditions are tight.

In the FLH-Model, the transaction cost for a firm i is obtained by $c_t^i = c_{0,t_p}^i + c_{1,t_p}^i.FL_t$. To estimate the transaction cost of two stocks i and j for a given time period t_p, we will estimate the parameters c_{0,t_p}^i and c_{1,t_p}^i for stock i and c_{0,t_p}^j and c_{1,t_p}^j for stock j. If $c_{1,t_p}^i > c_{1,t_p}^j$, the transaction cost differential between stock i and stock j will increase with the TED spread.

Table 1.3 presents the average value of parameters c_1 for size and realized volatility decile portfolios. The values for c_1 decrease with size and increase with volatility, supporting the flight-to-quality condition. The other columns in Table 1.3 confirm that the transaction cost decreases with size and increases with volatility in absolute terms, but that it increases with size and decreases with volatility in percentage, which is consistent with what was apparent in the time-series evolution of the size and volatility portfolios.

Table 1.3 here.

d. Breaking down the transaction cost into its fixed and time-varying parts

In Figure 1.9, we separate the average transaction cost for all firms into its fixed part and its time-varying part. We plot the time series of the c_0, which is fixed for a year, and of $c_1 \cdot FL_t$ which varies with the level of the TED spread. We see again the downtrend in the fixed part and the time-varying that mimics a scaled version of the TED time series. When funding conditions are really tight, as during the 2008 crisis, the part of the transaction cost that depends on the funding liquidity can be more important than the other part.

Figure 1.9 here.

1.4.3 Transaction Costs and other Financial Risk Measures

In this section, we summarize the main results associated with two risk measures that potentially affect the magnitude of the transaction cost. We estimate the time-varying part of the transaction cost $c_1^i \cdot FR_t$, where FR_t is in turn the VIX and the Weller [2019] measure of tail risk. We report the average transaction costs for the anomalies that are most impacted by the financial risk, that is anomalies related to size and realized volatility, as well as their dynamics.

a. Transaction Costs and the VIX

The Chicago Board Option Exchanges (CBOE) Market Volatility Index, or VIX is a popular measure of the stock market's expectation of volatility implied by S&P 500 index options. The VIX is often referred to as a "fear index" or the "fear gauge"(Whaley [2000]) for asset markets. A high value of the VIX is interpreted by investors as a potential sharp move of the market, either upward or downward, that is a higher expected volatility. Gromb and Vayanos [2002] and Brunnermeier and Pedersen [2009] predict that a higher market volatility tightens funding constraints of market makers and thereby reduces their liquidity-provision capacity. Nagel [2012] argues that when the VIX is high, market makers are financially constrained and therefore require a higher premium. More concretely, a higher market volatility makes stock prices move further away from their fundamental value, and therefore increase transaction costs.

Similarly to what we did for funding liquidity, we estimate the Hasbrouck model (H-Model) and the model with the VIX (VIXH-Model). Our sample covers the period from January 1990 to June 2018. In Table 1.4, we report the average round-trip transaction cost for anomalies that produce the largest differences between the two extreme decile portfolios, that is the size of the firm (measured by SIZE, PRICE and NOA) and the realized volatility of the form (REALVOL). The spread between the trading costs of two extreme deciles for the three size-anomaly portfolios are wider than for funding liquidity. For SIZE, when we take in account the VIX in the Hasbrouck model, the transaction cost of the small-firm portfolio (D1) increases in average by 660 basis points while the big-firm one (D10) is 70 basis points larger. Overall, the spread between D1 and D10 for the VIXH-Model is 10%. For realized volatility, the trading-cost differential between D10 and D1 for the VIXH-Model more than doubled with respect to the FLH model (16.6% instead of 8%). The statistical support for the VIXH-Model against the H model is again very strong for the larger-firm and the lower-volatility portfolios. The fact that VIX shocks include funding shocks and shocks from other sources may explain these larger spreads between extreme portfolios.

Table 1.4 and Figure 1.10 here.

Figure 1.10 show the dynamics of the transaction costs estimated from the two models (H-Model and VIXH-Model) for small and large firms and for high- and low-volatility portfolios. For the size-based anomaly portfolios, the patterns we uncovered with funding liquidity as a measure of financial risk remain the same both in terms of trend and large peaks, but the spreads between the H-Model and the VIXH-Model have widened considerably, as already indicated by the averages. For small firms, the differential in the beginning of the 90s is now close to 6%, while for large firms the peak around 2000 generates a spread o more than 2%. This increase in the trading-cost differential between the VIXH-Model and the H-Model is also present for the volatility portfolios. The VIX as a financial measure increases more the transaction costs of firms and produces relatively more peaks than funding liquidity. Flight to quality is also strongly supported when the VIX is used as the financial risk variable[13].

b. Transaction Costs and Tail Risk

Market participants and regulators can rely on two prominent measures of high-frequency tail risk developed by Bollerslev and Todorov [2011] and Weller [2019]. The first paper uses high-frequency intra-daily data and short maturity out-of-the-money options on the S&P 500 index to construct an Investors Fears index. The second paper stresses the potential limitations imposed by the rarity of liquid, deep out-of-the-money options and proposes a new methodology that relies on the cross-section of bid-ask spreads. In terms of risk factors associated with extreme events, the second measure captures the aggregate economic shocks and the potential systemic threats underlying the cross-section of realized stock returns, while the first measure picks up the risk factors extracted from liquid options on the S&P 500 index.

In this section we estimate a model of transaction cost with tail risk (TRH-Model) using the measure proposed by Weller [2019]. Since it is an extreme risk factor extracted from high-frequency quote data for thousands of U.S. stocks, it seems particularly appropriate for

[13] See detailed results in Table 1.11 in the Appendix.

our analysis. The paper concentrates on the 2008 financial crisis and its aftermath so our trading-cost estimates will cover only the period from January 2008 to December 2014.

With respect to the relative importance of the VIX and tail risk to measure financial risk, Bollerslev et al. [2015] decompose the VIX into a jump tail risk component and normal-sized price fluctuations. They show that the compensation for jump tails risk makes up a larger part of the variance risk premium. Therefore, it will be interesting to measure the transaction cost associated with a tail risk measure. However, with the Weller [2019] measure, the scope of our analysis will be mainly focused on the 2008 financial crisis.

Table 1.1 and Table 1.5 here.

In Table 1.1, the average transaction cost for the H-Model is estimated at 176 basis points and the addition of tail risk adds only 20 basis points to the trading cost. Table 1.5 features the trading costs for the size-related and volatility decile portfolios. The spread between the smallest-firm portfolio (D1) and the largest-firm portfolio (D1) for SIZE is 190 basis points for the H-Model and is increasing by 20 basis points for the TRH-Model. However, the Bayes factor shows strong evidence for the model with tail risk with 98% for the large firms and 72% for small firms. The volatility spread is larger with a trading cost of 50 basis points for the low-volatility portfolio (D1) and 500 basis points for the high-volatility portfolio, without the tail risk. Including the latter adds 40 basis points to the spread. Again the Bayes factor is very supportive of the model with tail risk. Interestingly, the lowest support occurs for the two extreme portfolios.

Figure 1.11 here.

After the 2008-2009 crisis, as shown in Figure 1.11, the effective bid-ask spread subsided quickly for size and volatility. The patterns for both anomalies are similar to what was observed with the VIX in the later part of the sample. Finally, flight to quality is also supported with tail risk as the financial risk variable[14].

1.5 After-trading-cost Performance of Anomalies

In this section we will evaluate the effect of accounting for transaction costs on the performance of so-called anomaly strategies, which consist in building long-short portfolios by

[14] See detailed results in Table 1.12 in the Appendix.

going long on the stocks in the highest decile and short on the stocks in the lowest decile or inversely, depending on the anomaly. For example, for the volatility anomaly, the portfolio is obtained by going long on the stocks in the highest decile and short on those in the lowest decile. For size, it is the reverse. The portfolio is long on stocks in the lowest decile and short on stocks in the highest decile.

Each month, stocks are ranked based on the value of the anomaly variable and placed accordingly in one of the decile portfolios. Given this way of proceeding, each month or each year depending on the trading frequency, the stocks included in a given decile are not necessarily the same as in the previous month or year. Therefore, to stay exposed to the anomaly, the long-short portfolio needs to be rebalanced and transaction costs are incurred.

1.5.1 Gross Returns and Net Returns of Anomalies

To compute the alpha of a long-short strategy we need the returns of the corresponding portfolio. The gross returns of a portfolio are obtained by computing an average of the individual stock returns using either an equal weight or their respective capitalization values. The net return of a portfolio will then be equal to the gross return minus the transaction cost of the portfolio.

To compute the net returns of our portfolios, we proceed as Lesmond et al. [2004] or Brandt et al. [2009]. For an anomaly-based portfolio, transaction costs are incurred only when the portfolio is rebalanced. For example, if we compute the monthly returns of a portfolio that is rebalanced once a year say in June, the net return will be equal to the gross return for all months except June. For the latter, we will subtract the transaction costs associated with the rebalancing from the gross return. However if we rebalance monthly, we need to take the transaction costs out of the monthly gross returns of the portfolio.

We will limit ourselves to assessing the performance of monthly rebalanced long-short strategies since this is where transaction costs will matter the most. To explain how to compute the rebalancing costs, we take the example of an equally-weighted portfolio P. It consists of a set A_{t-1} of n stocks $\{s_1, s_2, ..., s_n\}$ with weights $\frac{1}{n}$ at month $t-1$. At month t, the portfolio is rebalanced and consists of a set A_t of m stocks with weights $\frac{1}{m}$. Then the net return of portfolio P at time t is:

$$netR_t = \sum_{i=1}^{m} R_{t,z_i} - \text{Transaction costs,} \tag{1.14}$$

where R_{t,z_i} is the return of stock z_i at t, and,

$$
\begin{aligned}
\text{Transaction costs} = \frac{1}{n} & \sum_{v \in A_{t-1} \setminus A_t} \text{t-cost for selling } v \\
+\frac{1}{m} & \sum_{v \in A_t \setminus A_{t-1}} \text{t-cost for buying } v \\
+\mathbf{1}_{n>m}(\frac{1}{m} - \frac{1}{n}) & \sum_{v \in A_{t-1} \bigcap A_t} \text{t-cost for buying } v \\
+\mathbf{1}_{m>n}(\frac{1}{n} - \frac{1}{m}) & \sum_{v \in A_t \bigcap A_{t-1}} \text{t-cost for selling } v
\end{aligned} \tag{1.15}
$$

The first line of the above formula refers to the selling of the stocks that were in the set A_{t-1} and are not in the set A_t, while the second line accounts for the buying of the stocks that were not in the set A_{t-1} and are now in the set A_t. The two other lines are due to the reweighting of the stocks that remain in the portfolio from $t-1$ to t. If $n > m$ the weight of these common stocks will increase at time t and then we will need to buy more of these stocks, while we will sell them if $m > n$.

1.5.2 Performance of Long-short Strategies

The performance of these long-short anomaly-based strategies will be measured by the *alpha* of the portfolio with respect to the Fama-French 3-factor model. This choice is consistent with the fact that many anomalies were uncovered with this benchmark set of factors. It is also a choice that will lean towards a more generous assessment of the performance before accounting for trading costs. The *alpha* of the strategy portfolio is obtained as the intercept of the following regression:

$$
R_{it} - R_{ft} = \alpha_i + \beta_1.(R_{Mt} - R_{ft}) + \beta_2.SMB_t + \beta_3.HML_t + \epsilon_{it}, \tag{1.16}
$$

where R_{it} is the total return of the strategy portfolio i, R_{ft} is the risk-free rate of return, R_{Mt} is the total market portfolio return, $R_{it} - R_{ft}$ is the excess return of the strategy, $R_{Mt} - R_{ft}$ is the excess return on the market portfolio, SMB_t is the size premium (small minus big), HML_t is the value premium (high minus low), all evaluated in month t, and β_1, β_2, and β_3 denote the factor loadings of the strategy portfolio.

Table 1.6, Table 1.7, and Table 1.8 here.

Table 1.6, Table 1.7 and Table 1.8 report the alphas of the anomaly strategies that are rebalanced monthly, 17 overall. The three tables correspond to the three financial risk measures we considered. In each table, we report the alphas for the gross returns, the returns net of the H-Model transaction costs and the returns net of the transaction costs associated with the financial risk model. The staggered availability of the financial risk measures gives us the opportunity to assess the performance of the anomaly-based portfolios in different sample periods. The longest period is from January 1986 to June 2018 for funding liquidity, the intermediate one is from January 1990 to June 2018 for the VIX, and the shortest one for the tail risk measure from January 2008 to December 2014. This will put forward the robustness of the performance of the so-called anomalies and tell us if they produce in average profits across time periods.

We first look at the equally-weighted portfolios (first three columns of the tables). A remarkable fact is that 11 of the 17 strategies produce significantly positive gross returns over the three sample periods. For the remaining six strategies, five produce non-significantly-different-from-zero gross returns and one (ROME, return on market equity) significantly negative gross returns. The other remarkable fact is that out of the 11 profitable strategies only three remained significantly positive after applying the transaction costs either without or with the financial risk included. The first is earnings surprises (measured by Standardized Unexpected Earnings, SUE, defined in the Appendix). After accounting for trading costs with financial risk, the monthly net returns remain between 0.75% and 1.66%. The price per share (PRICE) delivers a solid 3 to 4% net return per month over the three samples after deducting trading costs that include financial risk. For the third one, industry momentum (INDMOM), the net returns remain positive (0.52% and 3.03%) for funding liquidity and tail risk but become significantly negative for the VIX.

For the value-weighted portfolios, the two main facts are still present. Ten strategies produce significantly positive gross returns and the same three anomalies (SUE, PRICE and INDMOM) are profitable for the H-Model. The unexpected earnings does not survive the inclusion of financial risk, and the industry momentum produces unrealistically high net returns with and without financial risk.

The strategies that we studied are based on the high-minus-low decile sort that is most commonly employed in academic studies. Novy-Marx and Velikov [2016] stresses that they significantly overstate the actual cost of trading these anomalies. First, because in practice, large institutional investment firms devote considerable resources to reduce the costs of executing trades. Second, because these strategies were designed ignoring trading costs, and therefore generate too much trading and too high trading costs.They propose three simple, rule-based methodologies to mitigate the incurred trading costs, in particular limiting the universe of traded stocks to the cheap-to-trade ones, and significantly reduce turnover without significantly reducing exposure to the anomaly. We keep the study of these practical refinements in the presence of time-varying trading costs due to financial risk for future research.

1.6 Robustness Check: the Transaction-cost Model of Lesmond, Ogden and Trzcinka (LOT) (1999)

The model proposed by Lesmond et al. [1999] is rooted in the adverse selection framework of Glosten and Milgrom [1985] and Kyle [1985]. In this literature, the marginal investor or the informed investor will trade on new information or accumulated information only if the trade leads to a profit net of transaction costs. In the model presented by Glosten and Milgrom [1985], the informed investor trades with the market-maker. His decision to trade or not to trade a security j depends not only on the bid and ask prices, B_{jt} and A_{jt} respectively, but also on the value $Z_{jt}(I_{jt})$ associated with the set of information I_{jt} he has. So he decides to:

$$\begin{aligned} \text{Buy if} \quad & Z_{jt} > A_{jt} \\ \text{Sell if} \quad & Z_{jt} < B_{jt} \\ \text{Do not trade if} \quad & B_{jt} < Z_{jt} < A_{jt} \end{aligned} \tag{1.17}$$

Before a security begins to reflect new information, a threshold (transaction cost) needs to be exceeded. For a security with a high transaction cost, it will be less likely for it to reflect new information than for a security with a low transaction cost. Therefore, a high transaction-cost security will count more zero-returns days than a low-transaction cost security.

1.6.1 Specification of the LOT Model

The model requires only the daily securities returns to estimate the transaction costs for any firm given a time period. The key feature in their model is the incidence of zero-return days.

According to the model, the observed returns are not the true ones. The true returns are the ones observed net of transactions costs. A zero-return day means that traders know that it would be non-profitable for them to trade after accounting for the transaction costs.

Lesmond et al. [1999] propose a limited-dependent variable (LDV) model of the relationship between the observed return R_{jt} and the true one R^*_{jt}. They assume the true return of a security j is given by price responses to both contemporaneous market return and firm-specific information through the following equation:

$$R^*_{jt} = \beta_j R_{mt} + \epsilon_{jt}, \tag{1.18}$$

where R_{mt} is the market return, β_j is the sensitivity of the true return to the market return and ϵ_{jt} captures the price response to firm-specific information. ϵ_{jt} is assumed to be normally distributed. The relationship between the observed return R_{jt} and the true one R^*_{jt} is given by:

$$\begin{aligned} R_{jt} = R^*_{jt} - \alpha_{1j} \text{ if } \quad & \quad R^*_{jt} < \alpha_{1j} \\ R_{jt} = 0 \text{ if } \quad & \quad \alpha_{1j} \leq R*_{jt} \leq \alpha_{2j} \\ R_{jt} = R^*_{jt} - \alpha_{2j} \text{ if } \quad & \quad R^*_{jt} > \alpha_{2j} \end{aligned} \tag{1.19}$$

where $\alpha_{1j} < 0$ represents the cost of selling and $\alpha_{2j} > 0$ the cost of buying. α_{1j} and α_{2j} are not supposed to be equal since studies like Berkowitz et al. [1988] and Huang and Stoll [1994] have provided evidence that the selling cost exceeds the buying cost.

Equation (1.17) describes the behaviour of the marginal investor. She makes her decisions based on the true returns and not on the observed returns. Trading occurs when the true return exceeds the transaction cost and continues as long as this condition is met. The price adjusts until the transaction cost exceeds the true return. When it happens, trading stops and a zero-return is observed. It means that the price effect of the new information is not enough to motivate the marginal investor to trade.

1.6.2 Estimation of the LOT Model

The resulting likelihood function of the econometric structure of the model has three components for the observed return: one for the decreases, one for the increases, and one for the zeros:

$$L = \prod_{t \in R_1} \frac{1}{\sigma_j} \phi_1(\zeta_t) \prod_{t \in R_2} \frac{1}{\sigma_j} \phi_2(\zeta_t) \prod_{t \in R_0} Pr(\text{no change})_t, \tag{1.20}$$

where R_1 and R_2 are the regions where the observed market returns are negative and positive respectively. R_0 is the region of zero returns. ϕ_1 and ϕ_2 refer to normal standard density functions for decreases and increases in the observed market returns, respectively. ζ_t is the standardized residual defined as $\zeta = \frac{\epsilon}{\sigma}$, where σ^2 is the estimated variance of the residuals using only the non-zero observed returns. $Pr(\text{no change})_t$ is the probability of a zero return. The likelihood can be rewritten as:

$$\begin{aligned} L(\alpha_{1j}, \alpha_{2j}, \beta_j, \sigma_j | R_{jt}, R_{mt}) = \prod_1 \frac{1}{\sigma_j} n \left[\frac{R_{jt} + \alpha_{1j} - \beta_j R_{mt}}{\sigma_j} \right] \\ \prod_0 \left[N \left(\frac{\alpha_{2j} - \beta_j R_{mt}}{\sigma_j} \right) - N \left(\frac{\alpha_{1j} - \beta_j R_{mt}}{\sigma_j} \right) \right] \\ \prod_2 \frac{1}{\sigma_j} n \left[\frac{R_{jt} + \alpha_{2j} - \beta_j R_{mt}}{\sigma_j} \right], \end{aligned} \tag{1.21}$$

where $n(.)$ is the standard normal density function and $N(.)$ is the standard normal cumulative distribution function. The logarithm of the likelihood function in equation (1.21) is given by:

$$\begin{aligned} \log L(\alpha_{1j}, \alpha_{2j}, \beta_j, \sigma_j | R_{jt}, R_{mt}) = \sum_1 \ln \frac{1}{(2\pi\sigma_j^2)^{1/2}} + \\ \sum_1 \frac{1}{2\sigma_j^2} (R_{jt} + \alpha_{1j} - \beta_j R_{mt})^2 + \\ \sum_2 \ln \frac{1}{(2\pi\sigma_j^2)^{1/2}} + \sum_2 \frac{1}{2\sigma_j^2} (R_{jt} + \alpha_{2j} - \beta_j R_{mt})^2 + \\ \sum_0 \ln \left(\left[N \left(\frac{\alpha_{2j} - \beta_j R_{mt}}{\sigma_j} \right) - N \left(\frac{\alpha_{1j} - \beta_j R_{mt}}{\sigma_j} \right) \right] \right). \end{aligned} \tag{1.22}$$

The parameters $\alpha_{1j}, \alpha_{2j}, \beta_j$ and σ_j are obtained by maximizing the log-likelihood function expressed in equation (1.22). The transaction cost is given by $\alpha_{2j} - \alpha_{1j}$.

A key element in this model is how to define the three regions R_0, R_1 and R_2. The region R_0 is the set of days where we have zero returns. In the original paper of Lesmond et al. [1999], R_1 and R_2 are defined based on R_m, the market return. So R_1 is the set of days where R_m is negative and R_2 is the set of days where R_m is positive. However, Goyenko et al. [2009], who is concerned with market liquidity of individual securities, defines R_1 and R_2 based on R_j, the daily return of firm j. So R_1 is the set of days where R_j is negative and R_2 is the set of days where R_j is positive.

1.6.3 The LOT Model with Funding Liquidity

a. Model Specification

We extend the model of Lesmond et al. [1999] by making the transaction costs depend on the funding liquidity variable, the TED spread. In the original model, given a time period (a month, a quarter or a year), the transaction costs for a security are estimated by using the daily returns of the security and the daily market returns. For the given time period, the marginal investor will compare true returns to two thresholds (the selling transaction cost and the buying transaction cost) to decide if she will trade or not. The two thresholds are assumed to be constant during the time period. We want to relax this assumption and account for the fact that funding conditions may change during the time period considered.

To make the transaction cost time-varying, we write it as an affine function of the TED spread. Like in the original model, we assume that the true return of a security j is given by price responses to both contemporaneous market return and firm-specific information through the same equation as before:

$$R^*_{jt} = \beta_j R_{mt} + \epsilon_{jt}, \tag{1.23}$$

The relationship between the observed return R_{jt} and the true one R^*_{jt} is now given by:

$$\begin{aligned} R_{jt} = R^*_{jt} - \alpha_{1jt} \quad & \text{if} \quad R^*_{jt} < \alpha_{1jt} \\ R_{jt} = 0 \quad & \text{if} \quad \alpha_{1jt} \le R^*_{jt} \le \alpha_{2jt} \\ R_{jt} = R^*_{jt} - \alpha_{2jt} \quad & \text{if} \quad R^*_{jt} > \alpha_{2j} \end{aligned} \tag{1.24}$$

where α_{1jt} and α_{2jt} are the costs for selling and buying respectively and are given by:

$$\begin{aligned} \alpha_{1jt} &= \alpha_{01j} + \alpha_{1j} FL_t \\ \alpha_{2jt} &= \alpha_{02j} + \alpha_{2j} FL_t, \end{aligned} \tag{1.25}$$

where FL_t is the funding liquidity factor.

With this specification, each day, the marginal trader revisits her decisions as the two thresholds α_{1jt} and α_{2jt} are changing every day.

b. Model Estimation and Inference

The log likelihood function is now given by:

$$\begin{aligned}\log L(\alpha_{01j}, \alpha_{1j}, \alpha_{02j}, \alpha_{2j}, \beta_j, \sigma_j | R_{jt}, R_{mt}) = \sum_1 \ln \frac{1}{(2\pi\sigma_j^2)^{1/2}} + \\ \sum_1 \frac{1}{2\sigma_j^2}(R_{jt} + \alpha_{1FLjt} - \beta_j R_{mt})^2 + \\ \sum_2 \ln \frac{1}{(2\pi\sigma_j^2)^{1/2}} + \\ \sum_2 \frac{1}{2\sigma_j^2}(R_{jt} + \alpha_{2FLjt} - \beta_j R_{mt})^2 + \\ \sum_0 \ln \left(\left[N\left(\frac{\alpha_{2FLjt} - \beta_j R_{mt}}{\sigma_j}\right) - N\left(\frac{\alpha_{1FLjt} - \beta_j R_{mt}}{\sigma_j}\right) \right] \right),\end{aligned} \tag{1.26}$$

where $\alpha_{1FLjt} = \alpha_{01j} + \alpha_{1j} FL_t$ and $\alpha_{2FLjt} = \alpha_{02j} + \alpha_{2j} FL_t$.

The parameters $\alpha_{01j}, \alpha_{1j}, \alpha_{02j}, \alpha_{2j}, \beta_j$ and σ_j are obtained by maximizing the log-likelihood function in equation (1.26). The roundtrip transaction cost is given by $\alpha_{2FLjt} - \alpha_{1FLjt}$.

c. A Likelihood-ratio Test to compare the LOT Model to the LOT-FL Model

To see if data support the model with time-varying transaction costs, we perform a likelihood ratio test. The parameters of the LOT model with FL are $\alpha_{01j}, \alpha_{1j}, \alpha_{02j}, \alpha_{2j}, \beta_j$, and σ_j while the parameters of the LOT model are $\alpha_{1j}, \alpha_{2j}, \beta_j$, and σ_j. Since the LOT model is nested in the LOT model with FL (with α_{01j} and α_{02j} set to 0), the likelihood ratio test will test H_0 against H_1, where:

$$\begin{aligned} H_0 &: \alpha_{01j} = 0 \;\; \text{and} \;\; \alpha_{02j} = 0 \\ H_1 &: \alpha_{01j} \neq 0 \;\; \text{or} \;\; \alpha_{02j} \neq 0 \end{aligned} \tag{1.27}$$

The likelihood ratio statistic is given by:

$$LR = -2(l_1 - l_0) \tag{1.28}$$
$$LR \sim \chi(2)$$

where l_0 is the logarithm of the likelihood of the LOT model without funding liquidity and l_1 is the logarithm of the likelihood of the LOT model with funding liquidity.

1.6.4 Estimated Transaction Costs with the LOT and LOT-FL Models for Anomaly-based Portfolios

In Table 1.9, we report for each anomaly-based decile portfolio, the average selling transaction cost (α_1 and α_{1FL}), the average buying transaction cost (α_2 and α_{2FL}) and also the proportion Pr of firm-years for which the likelihood ratio test preferred the LOT model with funding liquidity to the LOT model without funding liquidity.

Table 1.9 here.

Berkowitz et al. [1988] and Huang and Stoll [1994] find that transaction costs are higher when selling than when buying. The results presented in Table 1.9 are consistent with their findings. This asymmetry, $|\alpha_1| \geq |\alpha_2|$ and $|\alpha_{1FL}| \geq |\alpha_{2FL}|$, is verified for each anomaly and each decile portfolio.

Table 1.9 also shows that the transaction costs are higher when estimated with funding liquidity, since in average $|\alpha_1| \leq |\alpha_{1FL}|$ and $|\alpha_2| \leq |\alpha_{2FL}|$ for each anomaly-based portfolio. Overall, adding funding liquidity to the LOT model increases the transaction costs in average by 10%. This may seem modest but what is important is that it increases when funding conditions are tight and adds a time-varying dimension to the estimation of transaction costs. Introducing the TED spread as a daily measure of funding liquidity we can follow daily the magnitude of the trading cost without having to rely on intra-daily trade and quotes data.

The patterns detected for the decile portfolios with the effective bid-ask spread approach are confirmed for all anomalies. For example, for all anomalies related to size, SIZE, NOA, PRICE and SHVOL, transaction costs decrease monotonically from the D1 portfolio to the D10 portfolio. The likelihood ratio test rejection rate is modest compared to what we obtained with the effective bid-ask approach of Hasbrouck [2009]. For the whole sample, it rejects the LOT model for 27% of firm-years at the 5% level , and 40% at the 10% level. We have to remember that, unlike the Hasbrouck model, the LOT model takes in account zero-trading days. When funding conditions worsen, the marginal investor may not trade and this will be reflected in the estimation of the transaction costs. To check this conjecture, ee run the regression of Z_t, the proportion of stocks for which we have zero trading, on the TED spread. The adjusted R^2 is 17% and the slope is significant at 1% significant level. By accounting for the incidence of zeros-trading days, the LOT model implicitly accounts for funding conditions.

1.6.5 Performance of Long-short Strategies with the LOT Models

We report in Table 1.10 the alphas associated with the 17 long-short strategies rebalanced monthly that we already studied in Tables 1.6, 1.7 and 1.8. We focus on funding liquidity only so the sample covers the period from January 1986 to June 2018. The gross returns do not change with respect to Table 1.6 where we recall 11 strategies produced significant positive alphas. As previously we considered equal-weighting and value-weighting when constructing the long-short portfolios.

Table 1.10 here.

The conclusions are clear and very similar to what we obtained with the Hasbrouck [2009] model with and without financial risk. Only two strategies survive the introduction of transaction costs: the industry momentum (INDMOM) with value-weighting only, and the price per share (PRICE) for both weighting schemes. This robustness check with another model to compute transaction costs shows that the anomaly profits disappear when transaction costs are taken into account. Of course, in computing this after-trading-cost performance, we strictly apply the rebalancing considered in the academic literature and do not allow for mitigating strategies employed by practitioners in financial institutions. But it shows clearly that the research about anomalies cannot be conducted without considering transaction costs and their dynamic behavior according to aggregate financial risks prevalent in the economy.

1.7 Conclusion

We have proposed extensions to the two main models used to compute transaction costs from daily returns on individual securities, the effective bid-ask spread model of Hasbrouck [2009] and the asymmetric bid-ask spread of Lesmond et al. [1999]. We introduced measures of financial risk in the estimation of these costs and showed how they can increase considerably in crisis times, whether these large shocks result from tight funding conditions, investors' fears signalled by the VIX or extreme events captured by tail risk. The estimation results are telling. Transaction costs feature large jumps at event times for portfolios built on firm size, realized volatility or momentum, but increase for many other firm characteristics. Our analysis also confirms the technological downward trend in transaction costs over the last 35 years or so, measures precisely how firm characteristics such as size and volatility affect the magnitude of the trading costs at a high frequency, and provides evidence about the flight-to-quality behavior that occurs in hard market times. Finally, the profitable long-short strategies that the academic literature has put forward based on some firm characteristics identify become either non-profitable or losing propositions. Over the many anomalies considered, only two related to price per share and industry momentum yield a solid profit. An important follow-up study will be to refine our analysis to introduce mitigating strategies used in practice to minimize trading costs by financial institutions and to follow precisely the dynamic performance of the strategies in the context of our time-varying transaction costs. This will make the academic literature on anomalies more relevant for practitioners.

Tables and Figures of Chapter 1

Table 1.1: **Summary statistics for the estimated transaction costs with each H-Model**

The results are based on a year-by-year analysis for three periods depending on the financial risk variable considered, January 1986 to June 2018 for the TED spread, January 1990 to June 2018 for the VIX and January 2008 to December 2004 for tail risk. Estimations are done using daily returns and daily equally-weighted market index returns. The transaction cost from the Hasbrouck [2009] model is given by $2\hat{c}$ and the transaction cost from the Hasbrouck model with financial risk is given by $2c\hat{_{FR}} = 2(c_0 + c_1.FR)$ where FR is in turn the TED spread, the VIX and the tail risk measure of Weller [2019].

Model	Period	Mean	Median	Std.Dev	Skewness	Kurtosis
H-Model	Jan 1986 - Jun 2018	0.0296	0.0194	0.0363	7.5121	193.9541
FLH-Model	Jan 1986 - Jun 2018	0.0335	0.0231	0.0380	8.0079	272.7709
H-Model	Jan 1990 - Jun 2018	0.0281	0.0184	0.0351	8.1921	232.1774
VIXH-Model	Jan 1990 - Jun 2018	0.0575	0.0261	0.2755	294.0054	1.1916e+05
H-Model	Jan 2008 - Dec 2014	0.0176	0.0126	0.0175	5.1678	95.8242
TRH-Model	Jan 2008 - Dec 2014	0.0196	0.0140	0.0195	4.7037	71.0814

Table 1.2: **Average H-Model and FLH-Model transaction costs for anomaly-based decile portfolios**

The results are based on a year-by-year analysis for the period January 1986 to June 2018. For each anomaly, we rank firms by using data on characteristics from CRSP and COMPUSTAT. Estimations are done using daily returns and daily equally-weighted market index returns. The transaction cost from the Hasbrouck [2009] model (H-Model) is given by $2\hat{c}$ and the transaction cost from the Hasbrouck model with funding liquidity (FLH-Model) is given by $2c_{\hat{FR}} = 2(c_0 + c_1.FL)$ where FL is turn the TED spread. Pr is the proportion of stock-years for which the Bayes factor preferred the FLH-Model to the H-Model.

Anomaly	T-cost	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
SIZE	$2 \times c$	0.065	0.044	0.035	0.029	0.025	0.022	0.019	0.017	0.015	0.013
	$2 \times c_{FL}$	0.070	0.049	0.040	0.033	0.029	0.026	0.023	0.021	0.018	0.016
	Pr	0.52	0.56	0.62	0.69	0.76	0.80	0.84	0.87	0.89	0.92
REALVOL	$2 \times c$	0.008	0.010	0.013	0.017	0.020	0.024	0.029	0.036	0.048	0.086
	$2 \times c_{FL}$	0.010	0.013	0.017	0.020	0.024	0.028	0.033	0.040	0.052	0.090
	Pr	0.80	0.79	0.81	0.80	0.79	0.78	0.76	0.73	0.67	0.52
INDMOM	$2 \times c$	0.033	0.032	0.032	0.032	0.027	0.023	0.024	0.027	0.029	0.033
	$2 \times c_{FL}$	0.037	0.036	0.036	0.036	0.031	0.027	0.028	0.031	0.033	0.037
	Pr	0.75	0.75	0.74	0.72	0.72	0.74	0.75	0.75	0.75	0.74
NISSA	$2 \times c$	0.026	0.022	0.024	0.031	0.032	0.027	0.028	0.031	0.031	0.028
	$2 \times c_{FL}$	0.030	0.026	0.028	0.036	0.037	0.030	0.032	0.034	0.035	0.032
	Pr	0.74	0.78	0.74	0.67	0.70	0.78	0.78	0.77	0.77	0.79
CISS	$2 \times c$	0.018	0.015	0.017	0.019	0.022	0.024	0.026	0.028	0.029	0.034
	$2 \times c_{FL}$	0.021	0.019	0.020	0.023	0.025	0.028	0.030	0.032	0.033	0.038
	Pr	0.77	0.77	0.79	0.81	0.80	0.80	0.80	0.79	0.79	0.75
MOM11	$2 \times c$	0.061	0.039	0.030	0.024	0.021	0.019	0.019	0.020	0.022	0.029
	$2 \times c_{FL}$	0.066	0.044	0.034	0.028	0.024	0.022	0.022	0.024	0.026	0.033
	Pr	0.63	0.71	0.73	0.74	0.75	0.76	0.78	0.79	0.80	0.81
MOM6	$2 \times c$	0.058	0.037	0.029	0.024	0.021	0.020	0.020	0.021	0.024	0.032
	$2 \times c_{FL}$	0.062	0.042	0.033	0.028	0.025	0.023	0.023	0.025	0.028	0.037
	Pr	0.65	0.72	0.74	0.74	0.75	0.76	0.77	0.79	0.80	0.78
LTREV	$2 \times c$	0.056	0.038	0.028	0.022	0.019	0.017	0.016	0.017	0.018	0.021
	$2 \times c_{FL}$	0.061	0.042	0.032	0.027	0.023	0.021	0.020	0.020	0.022	0.025
	Pr	0.64	0.71	0.74	0.76	0.77	0.78	0.79	0.81	0.83	0.85
STREV	$2 \times c$	0.050	0.033	0.026	0.023	0.023	0.022	0.021	0.023	0.026	0.039
	$2 \times c_{FL}$	0.055	0.037	0.030	0.026	0.027	0.026	0.025	0.027	0.031	0.044
	Pr	0.68	0.74	0.75	0.75	0.74	0.74	0.76	0.78	0.78	0.74
SEASON	$2 \times c$	0.044	0.030	0.025	0.022	0.020	0.019	0.019	0.021	0.024	0.032
	$2 \times c_{FL}$	0.049	0.035	0.028	0.025	0.023	0.023	0.023	0.024	0.027	0.036
	Pr	0.70	0.75	0.76	0.77	0.77	0.78	0.78	0.79	0.80	0.77

MOMREV	$2 \times c$	0.053	0.037	0.029	0.024	0.021	0.020	0.020	0.021	0.024	0.031
	$2 \times c_{FL}$	0.058	0.041	0.033	0.028	0.025	0.024	0.024	0.025	0.028	0.036
	Pr	0.66	0.72	0.74	0.75	0.75	0.76	0.77	0.79	0.79	0.78
NISSM	$2 \times c$	0.028	0.031	0.031	0.029	0.027	0.026	0.024	0.028	0.032	0.026
	$2 \times c_{FL}$	0.032	0.035	0.035	0.033	0.031	0.030	0.028	0.033	0.037	0.030
	Pr	0.79	0.77	0.77	0.78	0.78	0.76	0.73	0.69	0.70	0.74
INDRREV	$2 \times c$	0.050	0.033	0.027	0.023	0.022	0.021	0.021	0.023	0.027	0.039
	$2 \times c_{FL}$	0.055	0.038	0.031	0.027	0.025	0.024	0.025	0.027	0.031	0.044
	Pr	0.68	0.74	0.75	0.75	0.75	0.76	0.76	0.77	0.78	0.73
PRICE	$2 \times c$	0.066	0.040	0.030	0.023	0.019	0.017	0.016	0.015	0.014	0.012
	$2 \times c_{FL}$	0.070	0.044	0.034	0.027	0.023	0.020	0.020	0.019	0.017	0.015
	Pr	0.58	0.68	0.73	0.75	0.77	0.80	0.83	0.86	0.88	0.90
SHVOL	$2 \times c$	0.040	0.031	0.028	0.027	0.026	0.026	0.026	0.026	0.027	0.029
	$2 \times c_{FL}$	0.045	0.036	0.032	0.031	0.030	0.030	0.030	0.030	0.031	0.033
	Pr	0.55	0.63	0.69	0.74	0.77	0.80	0.81	0.82	0.84	0.84
IK	$2 \times c$	0.050	0.035	0.030	0.027	0.027	0.028	0.028	0.030	0.032	0.037
	$2 \times c_{FL}$	0.055	0.042	0.037	0.034	0.033	0.034	0.035	0.036	0.038	0.042
	Pr	0.63	0.72	0.75	0.78	0.79	0.79	0.79	0.79	0.79	0.76
IG	$2 \times c$	0.047	0.038	0.032	0.028	0.026	0.025	0.025	0.027	0.031	0.038
	$2 \times c_{FL}$	0.051	0.042	0.036	0.032	0.031	0.030	0.030	0.032	0.035	0.042
	Pr	0.65	0.72	0.76	0.79	0.81	0.82	0.82	0.80	0.78	0.71
NOA	$2 \times c$	0.052	0.053	0.044	0.037	0.032	0.028	0.024	0.020	0.017	0.014
	$2 \times c_{FL}$	0.057	0.057	0.049	0.042	0.037	0.032	0.028	0.025	0.021	0.018
	Pr	0.64	0.61	0.64	0.69	0.73	0.78	0.81	0.84	0.87	0.90
AG	$2 \times c$	0.050	0.040	0.033	0.028	0.025	0.025	0.024	0.025	0.027	0.032
	$2 \times c_{FL}$	0.052	0.043	0.037	0.033	0.030	0.029	0.029	0.030	0.032	0.036
	Pr	0.66	0.71	0.74	0.75	0.77	0.78	0.79	0.79	0.79	0.78
IA	$2 \times c$	0.049	0.040	0.033	0.031	0.029	0.027	0.027	0.028	0.029	0.032
	$2 \times c_{FL}$	0.053	0.044	0.039	0.037	0.034	0.033	0.032	0.033	0.034	0.037
	Pr	0.66	0.71	0.71	0.76	0.78	0.79	0.79	0.79	0.80	0.78
LEV	$2 \times c$	0.022	0.019	0.018	0.017	0.017	0.017	0.018	0.020	0.020	0.024
	$2 \times c_{FL}$	0.026	0.024	0.022	0.022	0.021	0.022	0.023	0.025	0.025	0.028
	Pr	0.86	0.88	0.88	0.88	0.87	0.87	0.85	0.83	0.83	0.72
ROAA	$2 \times c$	0.056	0.048	0.038	0.028	0.024	0.023	0.023	0.022	0.023	0.026
	$2 \times c_{FL}$	0.061	0.052	0.042	0.034	0.030	0.029	0.028	0.028	0.029	0.033
	Pr	0.64	0.68	0.71	0.72	0.75	0.78	0.80	0.81	0.82	0.81
SUE	$2 \times c$	0.031	0.030	0.030	0.030	0.031	0.030	0.028	0.029	0.029	0.029
	$2 \times c_{FL}$	0.035	0.034	0.034	0.034	0.035	0.033	0.032	0.033	0.033	0.034
	Pr	0.75	0.76	0.76	0.76	0.75	0.77	0.78	0.78	0.78	0.78

ROME	$2 \times c$	0.015	0.013	0.013	0.010	0.009	0.010	0.008	0.009	0.008	0.011
	$2 \times c_{FL}$	0.018	0.016	0.016	0.014	0.012	0.012	0.010	0.011	0.012	0.014
	Pr	0.95	0.94	0.90	0.90	0.94	0.87	0.92	0.90	0.97	0.96
ROBE	$2 \times c$	0.036	0.037	0.036	0.031	0.027	0.024	0.025	0.028	0.031	0.036
	$2 \times c_{FL}$	0.040	0.041	0.040	0.035	0.032	0.029	0.030	0.032	0.035	0.040
	Pr	0.72	0.72	0.73	0.75	0.77	0.80	0.80	0.79	0.76	0.73
SG	$2 \times c$	0.042	0.035	0.030	0.027	0.025	0.025	0.026	0.028	0.031	0.037
	$2 \times c_{FL}$	0.045	0.038	0.034	0.031	0.030	0.030	0.031	0.033	0.035	0.041
	Pr	0.69	0.73	0.75	0.77	0.78	0.78	0.79	0.79	0.77	0.73
GPROF	$2 \times c$	0.046	0.028	0.025	0.027	0.028	0.028	0.028	0.028	0.029	0.032
	$2 \times c_{FL}$	0.054	0.037	0.033	0.034	0.035	0.035	0.034	0.035	0.036	0.039
	Pr	0.69	0.76	0.78	0.79	0.79	0.79	0.79	0.79	0.79	0.76
GMARGINS	$2 \times c$	0.046	0.032	0.030	0.029	0.029	0.029	0.029	0.030	0.031	0.032
	$2 \times c_{FL}$	0.050	0.039	0.037	0.036	0.035	0.036	0.037	0.038	0.039	0.040
	Pr	0.69	0.74	0.76	0.76	0.77	0.77	0.78	0.78	0.78	0.77
FSCORE	$2 \times c$	0.045	0.046	0.026	0.039	0.028	0.036	0.026	0.030	0.028	0.025
	$2 \times c_{FL}$	0.048	0.057	0.035	0.043	0.033	0.040	0.031	0.035	0.032	0.030
	Pr	0.67	0.67	0.70	0.66	0.75	0.72	0.77	0.77	0.78	0.80
ATURNOVER	$2 \times c$	0.043	0.031	0.031	0.031	0.030	0.029	0.030	0.031	0.033	0.033
	$2 \times c_{FL}$	0.049	0.039	0.038	0.038	0.037	0.036	0.037	0.038	0.039	0.040
	Pr	0.68	0.76	0.76	0.78	0.79	0.79	0.77	0.76	0.74	0.73
SP	$2 \times c$	0.029	0.018	0.016	0.016	0.016	0.016	0.017	0.018	0.021	0.027
	$2 \times c_{FL}$	0.031	0.022	0.021	0.020	0.020	0.020	0.021	0.023	0.025	0.031
	Pr	0.76	0.89	0.90	0.90	0.90	0.89	0.89	0.86	0.84	0.78
ACC	$2 \times c$	0.049	0.037	0.031	0.027	0.025	0.026	0.027	0.029	0.031	0.035
	$2 \times c_{FL}$	0.052	0.042	0.036	0.032	0.030	0.031	0.032	0.034	0.036	0.041
	Pr	0.67	0.74	0.77	0.79	0.79	0.79	0.79	0.79	0.78	0.74
GLTNOA	$2 \times c$	0.027	0.037	0.045	0.049	0.043	0.036	0.029	0.024	0.020	0.016
	$2 \times c_{FL}$	0.031	0.040	0.047	0.051	0.046	0.039	0.032	0.027	0.024	0.020
	Pr	0.83	0.75	0.68	0.63	0.65	0.71	0.78	0.83	0.87	0.90

Table 1.3: **FLH-Model: Transaction costs and flight to quality**

The table features the average absolute change and the average change in percentage between the transaction costs estimated with the Hasbrouck [2009] model with funding liquidity (FLH-Model) and the Hasbrouck model (H-Model). For the two anomalies, size and volatility, we report the changes for the ten decile portfolios. We also compute the average parameter c_1 per portfolio since this parameter measures the sensitivity of the portfolio transaction cost to funding liquidity. We perform an ANOVA to test the difference of all these values across the deciles.

	Size			Volatility		
	Absolute change	Change in percentage	Parameter c_1	Absolute change	Change in percentage	Parameter c_1
Dec1	0.0046	09.01	0.5250	0.0026	39.27	0.2647
Dec2	0.0045	11.22	0.5407	0.0030	32.94	0.3377
Dec3	0.0044	13.55	0.5253	0.0034	29.28	0.3905
Dec4	0.0041	15.45	0.4945	0.0038	25.46	0.4345
Dec5	0.0041	17.32	0.4842	0.0041	22.93	0.4712
Dec6	0.0039	18.82	0.4613	0.0043	20.17	0.4998
Dec7	0.0037	20.65	0.4388	0.0044	16.95	0.5268
Dec8	0.0036	22.26	0.4242	0.0046	14.18	0.5604
Dec9	0.0034	24.52	0.4061	0.0046	10.66	0.5685
Dec10	0.0032	26.99	0.3887	0.0041	05.42	0.5205
Number of periods	390 months	390 months	33 years	395 months	395 months	33 years
Anova: F stat	56.87***	6.97***	5.21***	139.67***	14.34***	12.57***
Anova: DF Columns	9	9	9	9	9	9
Anova: DF Errors	3890	3890	320	3940	3940	320

Table 1.4: **Average H-Model and VIXH-Model transaction costs for anomaly-based decile portfolios**

The results are based on a year-by-year analysis for the period January 1990 to June 2018. For each anomaly, we rank firms by using data on characteristics from CRSP and COMPUSTAT. Estimations are done using daily returns and daily equally-weighted market index returns. The transaction cost from the Hasbrouck [2009] model (H-Model) is given by $2\hat{c}$ and the transaction cost from the Hasbrouck model with the VIX (VIXH-Model) is given by $2\hat{c_{VIX}} = 2(c_0 + c_1.VIX)$ where the VIX is the CBOE volatility index. Pr is the proportion of stock-years for which the Bayes factor preferred the VIXH-Model to the H-Model.

Anomaly	T-cost	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
SIZE	$2 \times c$	0.061	0.041	0.033	0.027	0.024	0.021	0.019	0.017	0.015	0.013
	$2 \times c_{vix}$	0.127	0.073	0.056	0.045	0.039	0.035	0.031	0.028	0.024	0.020
	Pr	0.56	0.58	0.63	0.71	0.77	0.81	0.85	0.88	0.90	0.92
REALVOL	$2 \times c$	0.007	0.010	0.013	0.016	0.019	0.023	0.028	0.035	0.046	0.085
	$2 \times c_{vix}$	0.016	0.018	0.022	0.028	0.035	0.041	0.050	0.063	0.088	0.182
	Pr	0.68	0.84	0.84	0.83	0.81	0.80	0.79	0.75	0.69	0.57
PRICE	$2 \times c$	0.065	0.040	0.029	0.023	0.019	0.017	0.016	0.015	0.014	0.012
	$2 \times c_{vix}$	0.117	0.065	0.048	0.037	0.030	0.027	0.026	0.025	0.022	0.020
	Pr	0.57	0.67	0.72	0.75	0.77	0.81	0.84	0.87	0.89	0.91
NOA	$2 \times c$	0.050	0.051	0.042	0.035	0.031	0.027	0.024	0.020	0.017	0.014
	$2 \times c_{vix}$	0.100	0.096	0.082	0.062	0.054	0.046	0.040	0.034	0.028	0.022
	Pr	0.63	0.60	0.64	0.69	0.73	0.77	0.81	0.84	0.86	0.89

Table 1.5: **Average H-Model and TRH-Model transaction costs for anomaly-based decile portfolios**

The results are based on a year-by-year analysis for the period January 2008 to December 2014. For each anomaly, we rank firms by using data on characteristics from CRSP and COMPUSTAT. Estimations are done using daily returns and daily equally-weighted market index returns. The transaction cost from the Hasbrouck [2009] model (H-Model) is given by $2\hat{c}$ and the transaction cost from the Hasbrouck model with tail risk (TRH-Model) is given by $2\hat{c_{TR}} = 2(c_0 + c_1.TR)$ where TR is the tail risk measure of Weller [2019]. Pr is the proportion of stock-years for which the Bayes factor preferred the TRH-Model to the H-Model.

Anomaly	T-cost	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
SIZE	$2 \times c$	0.029	0.025	0.020	0.017	0.016	0.015	0.014	0.013	0.011	0.010
	$2 \times c_{tr}$	0.032	0.027	0.023	0.019	0.018	0.016	0.015	0.014	0.013	0.011
	Pr	0.72	0.69	0.76	0.88	0.93	0.94	0.96	0.97	0.98	0.98
REALVOL	$2 \times c$	0.005	0.006	0.008	0.010	0.013	0.015	0.018	0.022	0.029	0.050
	$2 \times c_{tr}$	0.006	0.007	0.009	0.012	0.014	0.017	0.020	0.024	0.032	0.055
	Pr	0.77	0.95	0.96	0.95	0.94	0.92	0.90	0.88	0.80	0.66
PRICE	$2 \times c$	0.041	0.027	0.020	0.015	0.013	0.012	0.011	0.011	0.010	0.008
	$2 \times c_{tr}$	0.045	0.029	0.022	0.017	0.014	0.013	0.013	0.012	0.011	0.009
	Pr	0.67	0.80	0.86	0.88	0.88	0.93	0.96	0.97	0.97	0.96
NOA	$2 \times c$	0.030	0.022	0.019	0.018	0.018	0.018	0.018	0.020	0.021	0.025
	$2 \times c_{tr}$	0.033	0.024	0.021	0.019	0.019	0.019	0.020	0.021	0.023	0.027
	Pr	0.75	0.68	0.75	0.83	0.90	0.95	0.96	0.98	0.98	0.99

Table 1.6: **Alphas (in %) of anomaly portfolios with H-Model and FLH-Model transaction costs (January 1986 to June 2018)**

The performance of a strategy is measured by its *alpha*, that is the intercept in the regression $R_{it} - R_{ft} = \alpha_i + \beta_1.(R_{Mt} - R_{ft}) + \beta_2.SMB_t + \beta_3.HML_t + \epsilon_{it}$, where R_{it} is the total return of the strategy portfolio i, R_{ft} is the risk-free rate of return measured by the T-bill rate, R_{Mt} is the total market portfolio return, $R_{it} - R_{ft}$ is the excess return of the strategy, $R_{Mt} - R_{ft}$ is the excess return on the market portfolio, SMB_t is the size premium (small minus big), HML_t is the value premium (high minus low), all evaluated in month t, and β_1, β_2, and β_3 denote the factor loadings of the strategy portfolio. Fore each strategy, we report α_i (in %) and the t-statistic of α_i.

Anomaly		Equally-weighted			Value-weighted		
		Gross return	H-model net return	FLH-model net return	Gross return	H-model net return	FLH-model net return
SUE	α	3.2517	1.6931	1.4518	0.9994	0.2665	0.0609
	$t-stat$	29.2821	10.5375	8.2723	7.1524	1.7473	0.3832
ROME	α	-2.8221	-3.7628	-4.3198	-2.3839	-3.0008	-3.3130
	$t-stat$	-2.1316	-2.7934	-3.1012	-1.8434	-2.3153	-2.5474
ROBE	α	-0.1214	-2.1425	-2.4255	0.2081	-0.7306	-0.9695
	$t-stat$	-1.6185	-11.7253	-11.9663	1.6761	-4.9382	-6.1199
SG	α	1.1205	-0.9690	-1.2355	0.6165	-0.3328	-0.5601
	$t-stat$	12.4960	-5.0718	-5.8564	3.7614	-1.8766	-3.0269
INDMOM	α	6.5298	1.2190	0.5176	11.5458	9.2283	8.6702
	$t-stat$	27.9150	4.7539	1.9701	42.4822	37.1225	35.0551
NISSA	α	-0.5995	-3.4446	-3.8520	-1.5476	-2.7920	-3.1373
	$t-stat$	-1.5756	-7.2103	-7.6824	-2.8334	-4.9123	-5.5886
CISS	α	0.3558	-1.4369	-1.6814	-0.3843	-1.2496	-1.4588
	$t-stat$	2.3339	-8.5325	-9.7989	-2.3685	-7.7074	-8.9140
MOM11	α	0.3011	-1.8765	-2.1272	-0.5674	-2.2010	-2.5013
	$t-stat$	0.7493	-4.5521	-5.1355	-1.2039	-4.6816	-5.3354
MOM6	α	0.1776	-2.7262	-3.0573	-0.5267	-2.5299	-2.9105
	$t-stat$	0.4537	-6.7481	-7.5282	-1.1775	-5.6194	-6.4621
LTREV	α	0.7037	-0.3676	-0.4996	1.2410	0.4001	0.2530
	$t-stat$	2.6589	-1.4123	-1.9103	4.0903	1.3367	0.8459
STREV	α	1.5158	-5.2485	-5.9917	0.2079	-3.3930	-4.1116
	$t-stat$	4.4045	-14.7418	-16.5558	0.5400	-8.7486	-10.5432
SEASON	α	-0.2197	-6.2749	-6.9902	-0.2335	-3.3108	-3.9556
	$t-stat$	-1.4459	-28.6611	-30.5870	-0.9595	-13.3676	-15.8741
MOMREV	α	0.6458	-2.1519	-2.4715	0.9052	-0.9152	-1.2643

	$t-stat$	2.8007	-9.3010	-10.5302	2.9850	-3.0378	-4.1860
NISSM	α	0.3161	-0.4347	-0.5476	0.0038	-0.3833	-0.4878
	$t-stat$	2.5842	-3.2045	-3.8872	0.0226	-2.3058	-2.9299
INDRREV	α	1.7883	-5.0117	-5.7576	0.5892	-3.2097	-3.9385
	$t-stat$	5.9072	-15.9023	-17.8898	1.8932	-10.2826	-12.5174
PRICE	α	5.6287	4.2057	4.0685	3.7434	2.5590	2.4398
	$t-stat$	14.6890	10.8613	10.5051	9.6379	6.4799	6.1799
SHVOL	α	0.9253	-0.3485	-0.5169	0.5568	0.0065	-0.1141
	$t-stat$	4.9415	-1.7805	-2.5911	2.5173	0.0297	-0.5211

Table 1.7: **Alphas (in %) of anomaly portfolios with H-Model and VIXH-Model transaction costs (January 1990 to June 2018)**

The performance of a strategy is measured by its *alpha*, that is the intercept in the regression $R_{it} - R_{ft} = \alpha_i + \beta_1.(R_{Mt} - R_{ft}) + \beta_2.SMB_t + \beta_3.HML_t + \epsilon_{it}$, where R_{it} is the total return of the strategy portfolio i, R_{ft} is the risk-free rate of return measured by the T-bill rate, R_{Mt} is the total market portfolio return, $R_{it} - R_{ft}$ is the excess return of the strategy, $R_{Mt} - R_{ft}$ is the excess return on the market portfolio, SMB_t is the size premium (small minus big), HML_t is the value premium (high minus low), all evaluated in month t, and β_1, β_2, and β_3 denote the factor loadings of the strategy portfolio. Fore each strategy, we report α_i (in %) and the t-statistic of α_i.

Anomaly		Equally-weighted			Value-weighted		
		Gross return	H-model net return	VIXH-model net return	Gross return	H-model net return	VIXH-model net return
SUE	α	3.2958	1.7950	0.7462	1.1141	0.3707	-0.0329
	$t-stat$	26.3699	10.4759	2.9421	7.1218	2.1740	-0.1778
ROME	α	-2.8072	-3.4202	-3.8648	-2.3839	-3.0020	-3.4592
	$t-stat$	-2.1221	-2.6000	-2.9216	-1.8434	-2.3163	-2.6387
ROBE	α	-0.0637	-1.9694	-3.3986	0.3586	-0.5821	-1.1242
	$t-stat$	-0.7539	-10.1995	-9.9444	2.6715	-3.6379	-5.9856
SG	α	1.2441	-0.7246	-2.2423	0.7764	-0.1785	-0.7362
	$t-stat$	12.6324	-3.6944	-6.4935	4.3575	-0.9252	-3.4322
INDMOM	α	6.8468	1.8244	-1.6315	11.7178	9.3698	8.1188
	$t-stat$	26.2923	6.5843	-4.5961	37.8701	33.1305	29.4837
NISSA	α	-0.4640	-3.3061	-5.4066	-1.3637	-2.6112	-3.3031
	$t-stat$	-1.0097	-5.8341	-7.1862	-2.1001	-3.8666	-4.8394
CISS	α	0.3178	-1.3893	-2.9092	-0.2860	-1.1800	-1.6556
	$t-stat$	1.8832	-7.6601	-10.1148	-1.5724	-6.4931	-9.0265
MOM11	α	0.2683	-1.7973	-3.3984	-0.5006	-2.0964	-3.1082
	$t-stat$	0.5832	-3.8420	-6.8754	-0.9307	-3.9098	-5.8455
MOM6	α	0.2248	-2.5344	-4.5990	-0.4704	-2.4377	-3.6468
	$t-stat$	0.4975	-5.4731	-9.2192	-0.9202	-4.7456	-7.0889
LTREV	α	0.9778	-0.0654	-0.8295	1.4684	0.6403	0.1908
	$t-stat$	3.3311	-0.2263	-2.9437	4.3500	1.9224	0.5776
STREV	α	1.4555	-5.0586	-9.9017	0.2142	-3.3872	-5.5730
	$t-stat$	3.6924	-12.5758	-19.8006	0.4862	-7.6462	-12.0768
SEASON	α	-0.2513	-6.1687	-10.5633	-0.1744	-3.2695	-4.9307
	$t-stat$	-1.4625	-26.4309	-24.1479	-0.6303	-11.6275	-16.7049
MOMREV	α	0.7365	-1.9341	-3.9401	0.9459	-0.8462	-1.8599

	$t-stat$	2.8255	-7.4433	-13.9510	2.7600	-2.4846	-5.4236
NISSM	α	0.4294	-0.2843	-0.8186	0.0734	-0.3259	-0.5501
	$t-stat$	3.1595	-2.0199	-5.4800	0.3894	-1.7280	-2.9209
INDREV	α	1.7179	-4.8322	-9.7087	0.6386	-3.1453	-5.5196
	$t-stat$	4.9758	-13.7674	-21.3056	1.7927	-8.8586	-14.6278
PRICE	α	5.6736	4.2772	3.1320	3.9039	2.6982	1.8618
	$t-stat$	12.8930	9.6885	6.8407	8.7855	5.9743	3.9990
SHVOL	α	1.0826	-0.1064	-1.0781	0.7415	0.1860	-0.1630
	$t-stat$	5.3330	-0.5096	-4.8836	3.0763	0.7770	-0.6862

Table 1.8: **Alphas (in %) of anomaly portfolios with H-Model and TRH-Model transaction costs (January 2008 to December 2014)**

The performance of a strategy is measured by its *alpha*, that is the intercept in the regression $R_{it} - R_{ft} = \alpha_i + \beta_1.(R_{Mt} - R_{ft}) + \beta_2.SMB_t + \beta_3.HML_t + \epsilon_{it}$, where R_{it} is the total return of the strategy portfolio i, R_{ft} is the risk-free rate of return measured by the T-bill rate, R_{Mt} is the total market portfolio return, $R_{it} - R_{ft}$ is the excess return of the strategy, $R_{Mt} - R_{ft}$ is the excess return on the market portfolio, SMB_t is the size premium (small minus big), HML_t is the value premium (high minus low), all evaluated in month t, and β_1, β_2, and β_3 denote the factor loadings of the strategy portfolio. Fore each strategy, we report α_i (in %) and the t-statistic of α_i.

Anomaly		Equally-weighted			Value-weighted		
		Gross return	H-model net return	TRH-model net return	Gross return	H-model net return	TRH-model net return
SUE	α	2.8321	1.7965	1.6551	1.1324	0.5541	0.4596
	$t-stat$	9.9136	5.4630	4.8975	3.7153	1.7998	1.5080
ROME	α	-2.7298	-3.1724	-3.1756	-2.1687	-2.6130	-2.6275
	$t-stat$	-1.8236	-2.1107	-2.1148	-1.4213	-1.7009	-1.7056
ROBE	α	0.1422	-1.1151	-1.2866	0.8821	0.1174	-0.0272
	$t-stat$	1.1451	-4.3804	-4.5002	3.9116	0.4686	-0.1015
SG	α	1.2685	0.0034	-0.1535	0.3117	-0.4087	-0.5278
	$t-stat$	7.3604	0.0132	-0.5439	0.8962	-1.0814	-1.3490
INDMOM	α	6.7079	3.3604	3.0317	10.5354	8.7457	8.5201
	$t-stat$	18.5936	9.4217	8.5072	22.3396	19.8228	19.8788
NISSA	α	0.2370	-1.3888	-1.6196	0.3790	-0.3544	-0.4249
	$t-stat$	0.2170	-1.2522	-1.5039	0.5097	-0.4450	-0.6361
CISS	α	0.5291	-0.5454	-0.6527	-0.0767	-0.7758	-0.8410
	$t-stat$	1.7279	-1.7258	-2.0356	-0.1964	-2.0084	-2.1839
MOM11	α	0.4548	-0.8612	-0.9939	1.3525	0.1152	-0.0392
	$t-stat$	0.5302	-1.0173	-1.1853	1.1930	0.1028	-0.0357
MOM6	α	0.4875	-1.2975	-1.4736	1.5439	0.0555	-0.1228
	$t-stat$	0.6289	-1.6887	-1.9463	1.5767	0.0572	-0.1290
LTREV	α	0.9852	0.3206	0.2586	1.2366	0.6822	0.6420
	$t-stat$	1.8390	0.6041	0.4886	2.1014	1.1722	1.1121
STREV	α	0.8145	-3.4287	-3.8402	0.3394	-2.3186	-2.6372
	$t-stat$	1.4765	-6.2364	-7.1324	0.5395	-3.7014	-4.2684
SEASON	α	0.1128	-3.7015	-4.0615	0.1462	-2.1444	-2.3828
	$t-stat$	0.4030	-12.5227	-13.9695	0.3526	-5.2917	-5.6910
MOMREV	α	0.6709	-1.0164	-1.1692	1.3951	0.0872	-0.0315

	$t-stat$	1.6527	-2.5535	-2.9574	3.2864	0.2101	-0.0765
NISSM	α	0.6624	0.1051	0.0465	0.1210	-0.2582	-0.3155
	$t-stat$	2.7024	0.4186	0.1845	0.3338	-0.7180	-0.8862
INDRREV	α	0.9979	-3.2675	-3.6833	0.9026	-1.8736	-2.2234
	$t-stat$	1.9162	-6.2494	-7.2538	1.6028	-3.3454	-4.0503
PRICE	α	4.7939	4.1955	4.1254	3.3230	2.6940	2.6218
	$t-stat$	7.2969	6.3606	6.2455	4.7429	3.8102	3.6936
SHVOL	α	1.2311	0.4571	0.3732	0.3774	-0.0641	-0.1384
	$t-stat$	3.0428	1.1032	0.8946	0.8864	-0.1518	-0.3303

Table 1.9: **Average LOT selling and buying transaction costs for anomaly-based decile portfolios**

The results are based on a year-by-year analysis for the period January 2008 to December 2014. For each anomaly, we rank firms by using data on characteristics from CRSP and COMPUSTAT. For each anomaly and each firm, we estimate the parameters $\hat{\alpha}_{01j}, \hat{\alpha}_{1j}, \hat{\alpha}_{02j}$, and $\hat{\alpha}_{2j}$ by using daily returns and daily equally-weighted market index returns. The selling transaction cost is given by $\alpha_{1\hat{FL}j} = \hat{\alpha}_{01j} + \hat{\alpha}_{1j} * FL$ and the buying one by $\alpha_{2\hat{FL}j} = \hat{\alpha}_{02j} + \hat{\alpha}_{2j} * FL$, where FL is the TED spread. Pr is the proportion of firm-years for which the likelihood ratio test $H_0 : \alpha_{01j} = 0, \alpha_{02j} = 0$ versus $H_1 : \alpha_{01j} \neq 0, \alpha_{02j}$ rejects the constrained LOT model without liquidity.

Anomaly	T-cost	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
	α_1	-0.098	-0.050	-0.036	-0.029	-0.024	-0.022	-0.020	-0.018	-0.016	-0.014
	α_2	0.087	0.044	0.032	0.026	0.022	0.020	0.018	0.017	0.015	0.013
SIZE	α_{1FL}	-0.101	-0.050	-0.036	-0.029	-0.025	-0.022	-0.020	-0.018	-0.016	-0.014
	α_{2FL}	0.088	0.044	0.032	0.026	0.022	0.020	0.018	0.017	0.015	0.013
	Pr	0.280	0.291	0.289	0.287	0.286	0.272	0.259	0.255	0.254	0.262
	α_1	-0.011	-0.013	-0.016	-0.019	-0.024	-0.028	-0.034	-0.044	-0.056	-0.097
	α_2	0.011	0.012	0.015	0.018	0.022	0.026	0.031	0.039	0.050	0.081
REALVOL	α_{1FL}	-0.012	-0.014	-0.016	-0.020	-0.025	-0.029	-0.035	-0.045	-0.061	-0.121
	α_{2FL}	0.012	0.014	0.015	0.019	0.023	0.027	0.032	0.041	0.053	0.098
	Pr	0.265	0.258	0.258	0.255	0.255	0.258	0.267	0.276	0.292	0.312
	α_1	-0.039	-0.037	-0.036	-0.037	-0.030	-0.026	-0.027	-0.030	-0.033	-0.038
	α_2	0.035	0.033	0.032	0.034	0.028	0.024	0.025	0.027	0.030	0.034
INDMOM	α_{1FL}	-0.040	-0.038	-0.037	-0.038	-0.031	-0.027	-0.028	-0.031	-0.034	-0.039
	α_{2FL}	0.036	0.034	0.033	0.034	0.028	0.024	0.025	0.028	0.030	0.034
	Pr	0.276	0.274	0.269	0.268	0.287	0.271	0.282	0.278	0.266	0.263
	α_1	-0.030	-0.024	-0.024	-0.037	-0.043	-0.028	-0.031	-0.034	-0.035	-0.032
	α_2	0.027	0.022	0.022	0.034	0.039	0.025	0.028	0.030	0.031	0.029
NISSA	α_{1FL}	-0.030	-0.024	-0.024	-0.038	-0.043	-0.028	-0.032	-0.034	-0.035	-0.032
	α_{2FL}	0.027	0.022	0.022	0.034	0.039	0.025	0.028	0.030	0.031	0.028
	Pr	0.274	0.262	0.303	0.244	0.289	0.268	0.260	0.271	0.274	0.289
	α_1	-0.021	-0.016	-0.018	-0.021	-0.024	-0.028	-0.030	-0.032	-0.035	-0.043
	α_2	0.019	0.015	0.017	0.019	0.022	0.026	0.027	0.029	0.031	0.038
CISS	α_{1FL}	-0.021	-0.016	-0.018	-0.021	-0.024	-0.029	-0.031	-0.033	-0.036	-0.044
	α_{2FL}	0.020	0.015	0.017	0.019	0.022	0.026	0.028	0.029	0.032	0.039
	Pr	0.280	0.270	0.264	0.261	0.258	0.258	0.260	0.261	0.266	0.278
	α_1	-0.074	-0.046	-0.034	-0.027	-0.023	-0.021	-0.020	-0.022	-0.024	-0.032
	α_2	0.063	0.041	0.031	0.025	0.021	0.019	0.019	0.020	0.022	0.028
MOM11	α_{1FL}	-0.077	-0.047	-0.034	-0.028	-0.023	-0.021	-0.020	-0.022	-0.024	-0.032
	α_{2FL}	0.065	0.042	0.031	0.025	0.021	0.020	0.019	0.020	0.022	0.028
	Pr	0.305	0.289	0.285	0.284	0.279	0.268	0.263	0.258	0.251	0.257
	α_1	-0.069	-0.043	-0.033	-0.027	-0.024	-0.021	-0.021	-0.023	-0.026	-0.037
	α_2	0.060	0.039	0.030	0.025	0.022	0.020	0.020	0.021	0.024	0.032
MOM6	α_{1FL}	-0.073	-0.044	-0.034	-0.027	-0.024	-0.022	-0.021	-0.023	-0.026	-0.037
	α_{2FL}	0.062	0.039	0.031	0.025	0.022	0.020	0.020	0.021	0.024	0.032
	Pr	0.305	0.289	0.285	0.285	0.280	0.269	0.263	0.257	0.253	0.262

	α_1	-0.072	-0.043	-0.030	-0.024	-0.020	-0.018	-0.017	-0.017	-0.018	-0.021
	α_2	0.062	0.038	0.027	0.022	0.019	0.017	0.016	0.016	0.017	0.020
LTREV	α_{1FL}	-0.074	-0.044	-0.031	-0.024	-0.020	-0.018	-0.017	-0.017	-0.019	-0.022
	α_{2FL}	0.064	0.038	0.028	0.022	0.019	0.017	0.016	0.016	0.018	0.020
	Pr	0.279	0.274	0.268	0.267	0.268	0.266	0.260	0.264	0.259	0.259
	α_1	-0.058	-0.035	-0.027	-0.026	-0.030	-0.028	-0.025	-0.026	-0.028	-0.044
	α_2	0.050	0.031	0.025	0.024	0.028	0.026	0.023	0.024	0.025	0.038
STREV	α_{1FL}	-0.060	-0.035	-0.028	-0.026	-0.030	-0.029	-0.025	-0.026	-0.028	-0.045
	α_{2FL}	0.051	0.032	0.025	0.024	0.028	0.027	0.023	0.024	0.025	0.038
	Pr	0.295	0.280	0.278	0.282	0.277	0.275	0.266	0.264	0.259	0.267
	α_1	-0.053	-0.034	-0.027	-0.024	-0.022	-0.020	-0.021	-0.022	-0.025	-0.036
	α_2	0.047	0.031	0.024	0.022	0.020	0.019	0.019	0.020	0.023	0.032
SEASON	α_{1FL}	-0.055	-0.035	-0.027	-0.024	-0.022	-0.021	-0.021	-0.023	-0.026	-0.037
	α_{2FL}	0.048	0.031	0.025	0.022	0.020	0.019	0.019	0.021	0.023	0.033
	Pr	0.270	0.267	0.268	0.268	0.269	0.267	0.266	0.264	0.263	0.262
	α_1	-0.067	-0.043	-0.033	-0.027	-0.024	-0.022	-0.022	-0.023	-0.026	-0.036
	α_2	0.058	0.038	0.030	0.025	0.022	0.020	0.020	0.021	0.024	0.032
MOMREV	α_{1FL}	-0.069	-0.044	-0.034	-0.028	-0.024	-0.022	-0.022	-0.023	-0.026	-0.036
	α_{2FL}	0.059	0.039	0.030	0.025	0.022	0.020	0.020	0.021	0.024	0.033
	Pr	0.281	0.276	0.277	0.276	0.278	0.272	0.267	0.266	0.266	0.267
	α_1	-0.033	-0.035	-0.034	-0.031	-0.029	-0.027	-0.023	-0.038	-0.040	-0.029
	α_2	0.029	0.031	0.030	0.028	0.026	0.024	0.021	0.034	0.037	0.026
NISSM	α_{1FL}	-0.034	-0.036	-0.035	-0.032	-0.029	-0.027	-0.024	-0.039	-0.042	-0.030
	α_{2FL}	0.030	0.031	0.031	0.028	0.026	0.024	0.022	0.036	0.038	0.026
	Pr	0.290	0.278	0.273	0.261	0.267	0.266	0.278	0.293	0.264	0.273
	α_1	-0.058	-0.036	-0.030	-0.028	-0.026	-0.025	-0.025	-0.026	-0.029	-0.045
	α_2	0.050	0.032	0.027	0.026	0.024	0.023	0.023	0.024	0.026	0.039
INDRREV	α_{1FL}	-0.060	-0.037	-0.031	-0.028	-0.026	-0.025	-0.025	-0.026	-0.029	-0.045
	α_{2FL}	0.051	0.033	0.028	0.026	0.024	0.023	0.023	0.024	0.026	0.039
	Pr	0.293	0.278	0.276	0.279	0.278	0.274	0.270	0.264	0.262	0.267
	α_1	-0.058	-0.037	-0.028	-0.022	-0.019	-0.017	-0.017	-0.016	-0.014	-0.013
	α_2	0.047	0.032	0.025	0.021	0.018	0.016	0.016	0.015	0.013	0.012
PRICE	α_{1FL}	-0.060	-0.038	-0.029	-0.023	-0.019	-0.017	-0.017	-0.016	-0.014	-0.013
	α_{2FL}	0.048	0.033	0.025	0.021	0.018	0.016	0.016	0.015	0.014	0.012
	Pr	0.303	0.292	0.287	0.276	0.277	0.269	0.257	0.251	0.251	0.261
	α_1	-0.064	-0.039	-0.032	-0.029	-0.027	-0.026	-0.026	-0.027	-0.028	-0.030
	α_2	0.060	0.035	0.028	0.026	0.024	0.023	0.023	0.024	0.025	0.026
SHVOL	α_{1FL}	-0.066	-0.040	-0.033	-0.029	-0.027	-0.026	-0.026	-0.027	-0.028	-0.030
	α_{2FL}	0.060	0.036	0.029	0.026	0.024	0.023	0.024	0.024	0.025	0.027
	Pr	0.274	0.278	0.283	0.281	0.273	0.271	0.267	0.270	0.272	0.276
	α_1	-0.065	-0.041	-0.033	-0.029	-0.029	-0.029	-0.031	-0.033	-0.035	-0.042
	α_2	0.057	0.036	0.029	0.026	0.026	0.026	0.027	0.029	0.032	0.038
IK	α_{1FL}	-0.067	-0.041	-0.033	-0.030	-0.029	-0.030	-0.031	-0.033	-0.036	-0.043

	α_{2FL}	0.058	0.036	0.030	0.027	0.026	0.027	0.028	0.030	0.033	0.038
	Pr	0.289	0.274	0.271	0.266	0.264	0.265	0.261	0.260	0.259	0.263
	α_1	-0.058	-0.043	-0.034	-0.030	-0.028	-0.027	-0.027	-0.029	-0.033	-0.044
	α_2	0.051	0.038	0.031	0.027	0.025	0.024	0.024	0.026	0.030	0.039
IG	α_{1FL}	-0.060	-0.043	-0.035	-0.030	-0.028	-0.027	-0.027	-0.029	-0.034	-0.044
	α_{2FL}	0.052	0.038	0.031	0.027	0.025	0.024	0.024	0.026	0.031	0.039
	Pr	0.282	0.277	0.268	0.267	0.263	0.260	0.259	0.265	0.263	0.271
	α_1	-0.071	-0.070	-0.052	-0.039	-0.032	-0.026	-0.023	-0.020	-0.017	-0.014
	α_2	0.062	0.061	0.046	0.034	0.029	0.024	0.021	0.019	0.016	0.014
NOA	α_{1FL}	-0.073	-0.071	-0.053	-0.039	-0.033	-0.027	-0.023	-0.020	-0.018	-0.014
	α_{2FL}	0.064	0.062	0.046	0.034	0.029	0.024	0.021	0.019	0.016	0.014
	Pr	0.267	0.274	0.274	0.270	0.268	0.265	0.262	0.261	0.267	0.277
	α_1	-0.060	-0.046	-0.038	-0.032	-0.028	-0.026	-0.026	-0.027	-0.030	-0.036
	α_2	0.051	0.041	0.034	0.029	0.025	0.024	0.024	0.025	0.027	0.031
AG	α_{1FL}	-0.061	-0.047	-0.039	-0.032	-0.028	-0.027	-0.026	-0.027	-0.030	-0.036
	α_{2FL}	0.053	0.041	0.035	0.029	0.026	0.024	0.024	0.025	0.027	0.032
	Pr	0.288	0.278	0.275	0.271	0.267	0.264	0.262	0.259	0.259	0.265
	α_1	-0.061	-0.047	-0.040	-0.037	-0.033	-0.030	-0.029	-0.030	-0.031	-0.036
	α_2	0.053	0.041	0.036	0.033	0.030	0.027	0.026	0.027	0.028	0.033
IA	α_{1FL}	-0.063	-0.048	-0.041	-0.037	-0.033	-0.030	-0.030	-0.030	-0.031	-0.037
	α_{2FL}	0.055	0.042	0.036	0.033	0.030	0.027	0.027	0.027	0.029	0.033
	Pr	0.283	0.267	0.276	0.263	0.257	0.256	0.257	0.261	0.261	0.272
	α_1	-0.025	-0.021	-0.019	-0.019	-0.018	-0.019	-0.019	-0.021	-0.020	-0.024
	α_2	0.022	0.019	0.018	0.017	0.017	0.017	0.018	0.020	0.019	0.022
LEV	α_{1FL}	-0.025	-0.021	-0.020	-0.019	-0.019	-0.019	-0.020	-0.022	-0.021	-0.025
	α_{2FL}	0.022	0.019	0.018	0.017	0.017	0.017	0.018	0.020	0.019	0.023
	Pr	0.288	0.279	0.282	0.290	0.300	0.310	0.332	0.352	0.363	0.371
	α_1	-0.068	-0.059	-0.043	-0.030	-0.025	-0.024	-0.024	-0.024	-0.024	-0.029
	α_2	0.058	0.052	0.038	0.027	0.023	0.022	0.022	0.022	0.022	0.026
ROAA	α_{1FL}	-0.070	-0.060	-0.044	-0.030	-0.025	-0.025	-0.024	-0.024	-0.025	-0.029
	α_{2FL}	0.060	0.052	0.039	0.027	0.023	0.022	0.022	0.022	0.023	0.027
	Pr	0.288	0.286	0.291	0.289	0.276	0.266	0.252	0.245	0.242	0.246
	α_1	-0.032	-0.032	-0.032	-0.033	-0.034	-0.037	-0.031	-0.032	-0.032	-0.032
	α_2	0.029	0.029	0.029	0.029	0.031	0.034	0.027	0.028	0.029	0.028
SUE	α_{1FL}	-0.033	-0.032	-0.032	-0.033	-0.035	-0.038	-0.031	-0.032	-0.033	-0.033
	α_{2FL}	0.030	0.029	0.029	0.030	0.031	0.034	0.027	0.028	0.029	0.029
	Pr	0.287	0.279	0.276	0.274	0.267	0.261	0.262	0.257	0.257	0.254
	α_1	-0.017	-0.014	-0.014	-0.011	-0.010	-0.011	-0.008	-0.010	-0.009	-0.012
	α_2	0.016	0.014	0.014	0.011	0.010	0.011	0.009	0.010	0.009	0.012
ROME	α_{1FL}	-0.021	-0.013	-0.014	-0.011	-0.010	-0.012	-0.008	-0.010	-0.008	-0.012
	α_{2FL}	0.016	0.013	0.013	0.011	0.010	0.011	0.008	0.010	0.008	0.011
	Pr	0.489	0.322	0.289	0.254	0.341	0.242	0.156	0.276	0.184	0.446
	α_1	-0.040	-0.042	-0.042	-0.036	-0.030	-0.026	-0.027	-0.030	-0.035	-0.040

ROBE	α_2	0.036	0.038	0.037	0.032	0.027	0.024	0.024	0.027	0.031	0.036
	α_{1FL}	-0.041	-0.042	-0.042	-0.036	-0.030	-0.027	-0.027	-0.031	-0.036	-0.041
	α_{2FL}	0.036	0.038	0.038	0.032	0.027	0.024	0.025	0.027	0.032	0.036
	Pr	0.281	0.278	0.274	0.268	0.264	0.261	0.259	0.258	0.267	0.275
SG	α_1	-0.051	-0.039	-0.032	-0.028	-0.026	-0.026	-0.028	-0.031	-0.035	-0.044
	α_2	0.045	0.035	0.029	0.026	0.024	0.024	0.025	0.028	0.031	0.039
	α_{1FL}	-0.052	-0.040	-0.033	-0.029	-0.026	-0.026	-0.028	-0.031	-0.035	-0.045
	α_{2FL}	0.045	0.036	0.029	0.026	0.024	0.024	0.026	0.028	0.031	0.039
	Pr	0.278	0.278	0.272	0.267	0.266	0.261	0.259	0.260	0.264	0.270
GPROF	α_1	-0.054	-0.032	-0.028	-0.029	-0.030	-0.029	-0.029	-0.029	-0.031	-0.034
	α_2	0.047	0.028	0.025	0.026	0.027	0.026	0.026	0.026	0.028	0.030
	α_{1FL}	-0.055	-0.032	-0.028	-0.030	-0.030	-0.030	-0.029	-0.030	-0.032	-0.035
	α_{2FL}	0.047	0.028	0.025	0.027	0.027	0.027	0.026	0.027	0.028	0.031
	Pr	0.280	0.300	0.280	0.273	0.265	0.263	0.249	0.248	0.249	0.252
GMARGINS	α_1	-0.057	-0.034	-0.031	-0.031	-0.030	-0.032	-0.033	-0.034	-0.036	-0.039
	α_2	0.049	0.031	0.028	0.028	0.027	0.028	0.030	0.030	0.032	0.035
	α_{1FL}	-0.058	-0.035	-0.032	-0.031	-0.031	-0.032	-0.034	-0.034	-0.036	-0.040
	α_{2FL}	0.050	0.031	0.028	0.028	0.028	0.029	0.030	0.031	0.033	0.036
	Pr	0.285	0.275	0.271	0.266	0.261	0.258	0.260	0.253	0.261	0.259
FSCORE	α_1	-0.050	-0.067	-0.026	-0.051	-0.030	-0.043	-0.029	-0.034	-0.030	-0.028
	α_2	0.044	0.059	0.023	0.047	0.026	0.039	0.026	0.030	0.027	0.026
	α_{1FL}	-0.051	-0.071	-0.027	-0.052	-0.030	-0.044	-0.029	-0.035	-0.030	-0.029
	α_{2FL}	0.044	0.062	0.024	0.048	0.027	0.040	0.026	0.031	0.027	0.026
	Pr	0.283	0.352	0.278	0.335	0.256	0.302	0.246	0.307	0.239	0.254
ATURNOVER	α_1	-0.055	-0.038	-0.036	-0.035	-0.033	-0.031	-0.032	-0.033	-0.036	-0.036
	α_2	0.049	0.034	0.032	0.031	0.029	0.028	0.029	0.030	0.032	0.032
	α_{1FL}	-0.057	-0.038	-0.036	-0.035	-0.033	-0.032	-0.033	-0.034	-0.037	-0.037
	α_{2FL}	0.050	0.034	0.032	0.031	0.030	0.028	0.029	0.030	0.033	0.034
	Pr	0.292	0.280	0.278	0.264	0.257	0.260	0.255	0.254	0.258	0.260
SP	α_1	-0.032	-0.020	-0.018	-0.017	-0.017	-0.017	-0.018	-0.020	-0.022	-0.029
	α_2	0.027	0.018	0.016	0.016	0.016	0.016	0.017	0.018	0.020	0.026
	α_{1FL}	-0.033	-0.020	-0.018	-0.018	-0.017	-0.018	-0.018	-0.020	-0.022	-0.029
	α_{2FL}	0.027	0.018	0.017	0.016	0.016	0.016	0.017	0.019	0.021	0.026
	Pr	0.325	0.299	0.308	0.297	0.304	0.307	0.304	0.312	0.319	0.335
ACC	α_1	-0.059	-0.044	-0.036	-0.031	-0.028	-0.029	-0.030	-0.031	-0.034	-0.040
	α_2	0.050	0.039	0.032	0.028	0.026	0.026	0.027	0.028	0.031	0.035
	α_{1FL}	-0.060	-0.045	-0.036	-0.032	-0.029	-0.029	-0.030	-0.032	-0.035	-0.041
	α_{2FL}	0.052	0.039	0.032	0.029	0.026	0.027	0.028	0.029	0.031	0.035
	Pr	0.281	0.271	0.264	0.261	0.263	0.260	0.261	0.256	0.258	0.267
GLTNOA	α_1	-0.026	-0.040	-0.055	-0.069	-0.053	-0.039	-0.029	-0.024	-0.020	-0.016
	α_2	0.023	0.035	0.048	0.061	0.047	0.034	0.026	0.022	0.019	0.016
	α_{1FL}	-0.027	-0.041	-0.056	-0.070	-0.054	-0.039	-0.030	-0.024	-0.020	-0.017
	α_{2FL}	0.024	0.035	0.049	0.064	0.047	0.035	0.027	0.022	0.019	0.016
	Pr	0.271	0.273	0.270	0.267	0.267	0.258	0.252	0.248	0.252	0.271

Table 1.10: **Alphas (in %) of anomaly portfolios with LOT-Model and LOT-Model plus funding liquidity (FLLOT-Model) transaction costs (January 1986 to June 2018)**

The performance of a strategy is measured by its *alpha*, that is the intercept in the regression $R_{it} - R_{ft} = \alpha_i + \beta_1.(R_{Mt} - R_{ft}) + \beta_2.SMB_t + \beta_3.HML_t + \epsilon_{it}$, where R_{it} is the total return of the strategy portfolio i, R_{ft} is the risk-free rate of return measured by the T-bill rate, R_{Mt} is the total market portfolio return, $R_{it} - R_{ft}$ is the excess return of the strategy, $R_{Mt} - R_{ft}$ is the excess return on the market portfolio, SMB_t is the size premium (small minus big), HML_t is the value premium (high minus low), all evaluated in month t, and β_1, β_2, and β_3 denote the factor loadings of the strategy portfolio. Fore each strategy, we report α_i (in %) and the t-statistic of α_i.

Anomaly		Equally-weighted			Value-weighted		
		Gross return	LOT-model net return	FLLOT-model net return	Gross return	LOT-model net return	FLLOT-model net return
SUE	α	3.2517	0.0603	-0.0307	0.9994	-0.4550	-0.5039
	$t-stat$	29.2821	0.2150	-0.1063	7.1524	-2.4979	-2.7294
ROME	α	-2.8221	-4.7111	-4.8136	-2.3839	-3.6047	-3.6726
	$t-stat$	-2.1316	-6.3569	-6.8452	-1.8434	-2.7373	-2.7490
ROBE	α	-0.1214	-4.5070	-4.6240	0.2081	-1.6538	-1.7113
	$t-stat$	-1.6185	-11.1444	-11.0969	1.6761	-8.4804	-8.5325
SG	α	1.1205	-3.6193	-3.7339	0.6165	-1.2681	-1.3131
	$t-stat$	12.4960	-8.1186	-8.1738	3.7614	-5.8950	-5.9697
INDMOM	α	6.5298	-4.4149	-4.5747	11.5458	6.8539	6.7782
	$t-stat$	27.9150	-10.6301	-10.9413	42.4822	29.3454	29.5940
NISSA	α	-0.5995	-1.9456	-2.0345	-1.5476	-3.9929	-3.9762
	$t-stat$	-1.5756	-4.3023	-4.4092	-2.8334	-6.8944	-6.9507
CISS	α	0.3558	-3.5164	-3.5920	-0.3843	-1.9961	-2.0139
	$t-stat$	2.3339	-16.3614	-16.3611	-2.3685	-12.2001	-12.2826
MOM11	α	0.3011	-4.2049	-4.2801	-0.5674	-3.5088	-3.5591
	$t-stat$	0.7493	-9.4302	-9.4996	-1.2039	-7.4554	-7.6498
MOM6	α	0.1776	-5.9636	-6.0500	-0.5267	-4.2250	-4.2799
	$t-stat$	0.4537	-12.9788	-12.9999	-1.1775	-9.2901	-9.5169
LTREV	α	0.7037	-1.4050	-1.4431	1.2410	-0.1984	0.2085
	$t-stat$	2.6589	-5.3957	-5.5308	4.0903	-0.6742	-0.7111
STREV	α	1.5158	-12.8639	-13.0706	0.2079	-6.9982	-7.1261
	$t-stat$	4.4045	-24.4086	-24.6537	0.5400	-17.5717	-18.0859
SEASON	α	-0.2197	-13.3912	-13.5712	-0.2335	-6.3514	-6.4019
	$t-stat$	-1.4459	-28.5474	-28.1470	-0.9595	-24.5211	-24.5389

MOMREV	α	0.6458	-5.4137	-5.5033	0.9052	-2.4739	-2.4947
	$t-stat$	2.8007	-18.3619	-18.4405	2.9850	-8.1608	-8.2426
NISSM	α	0.3161	-1.1188	-1.1373	0.0038	-0.6923	-0.7042
	$t-stat$	2.5842	-7.7965	-7.9062	0.0226	-4.1542	-4.2425
INDRREV	α	1.7883	-12.6605	-12.8692	0.5892	-7.0000	-7.1294
	$t-stat$	5.9072	-25.2872	-25.5190	1.8932	-21.5402	-22.3322
PRICE	α	5.6287	3.6241	3.5562	3.7434	1.9452	1.8597
	$t-stat$	14.6890	9.2371	9.0673	9.6379	4.9088	4.6812
SHVOL	α	0.9253	-1,9214	-1.9530	0.5568	-0.4316	-0.4455
	$t-stat$	4.9415	-8.2908	-8.3237	2.5173	-2.0006	-2.0732

Figure 1.2: **Transaction costs increase from H-Model to FLH-Model and TED spread**

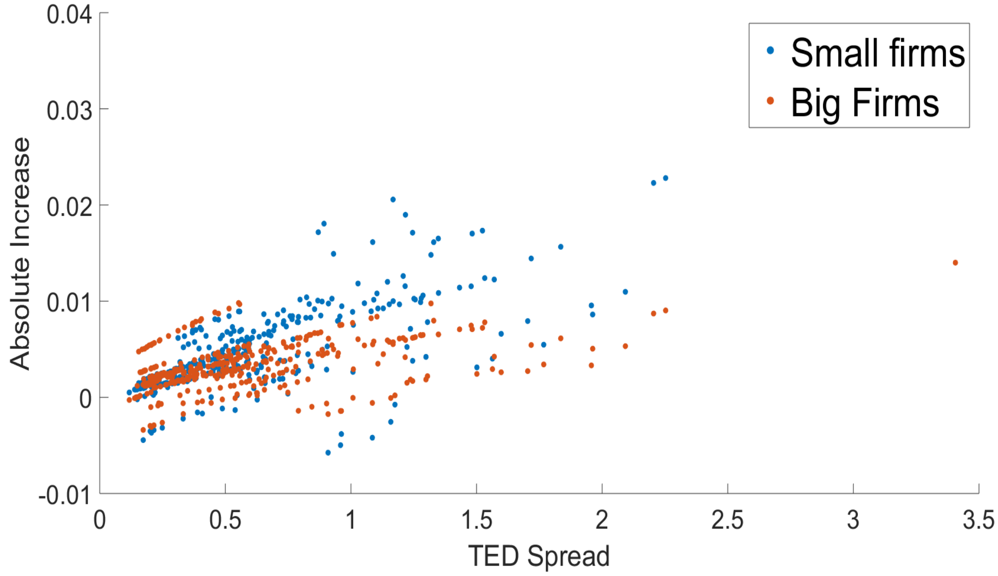


Figure 1.3: **Dynamics of transaction costs from H-Model and FLH-Model for COCA COLA CO**

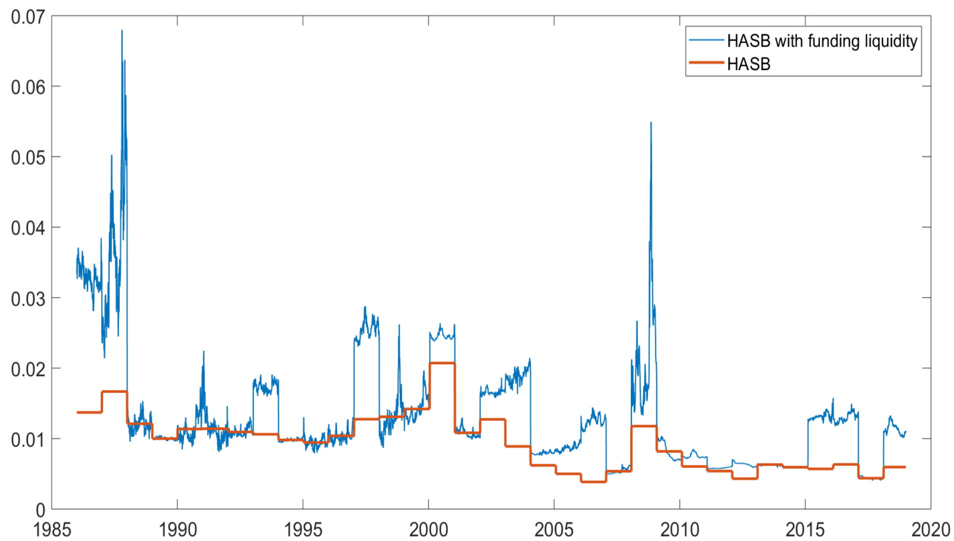


Figure 1.4: **Dynamics of transaction costs from H-Model and FLH-Model for ROCKY MOUNTAIN CHOCOLATE FACTORY**

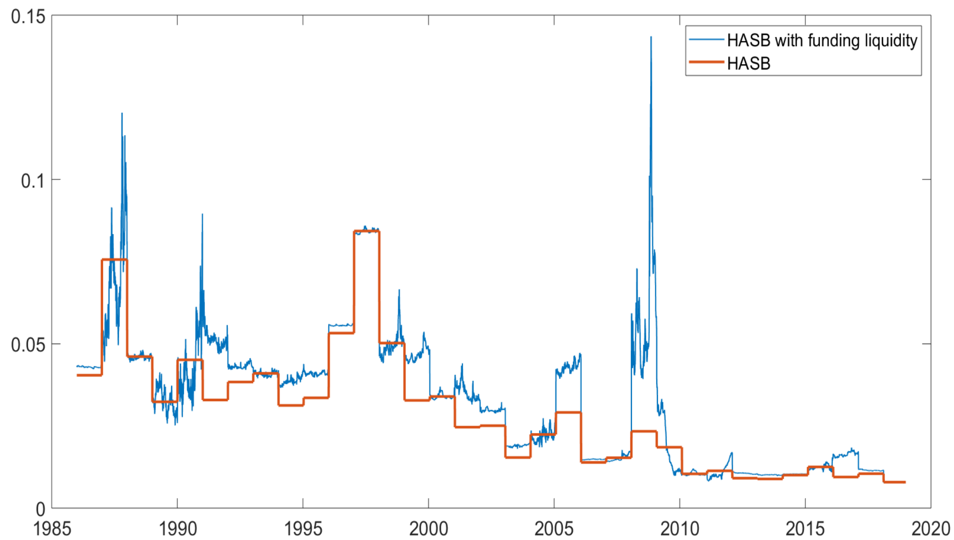


Figure 1.5: **Dynamics of transaction costs from H-Model and FLH-Model for small firms**

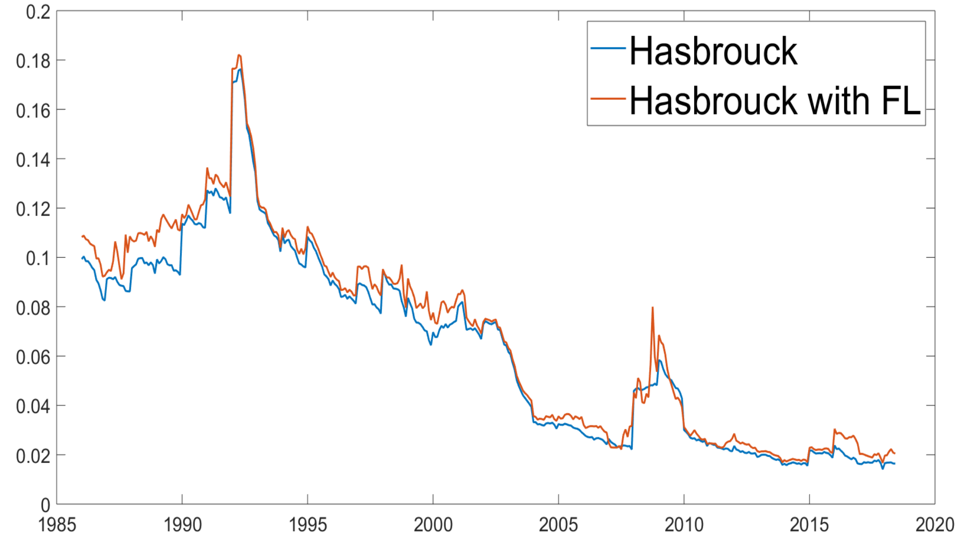


Figure 1.6: **Dynamics of transaction costs from H-Model and FLH-Model for big firms**

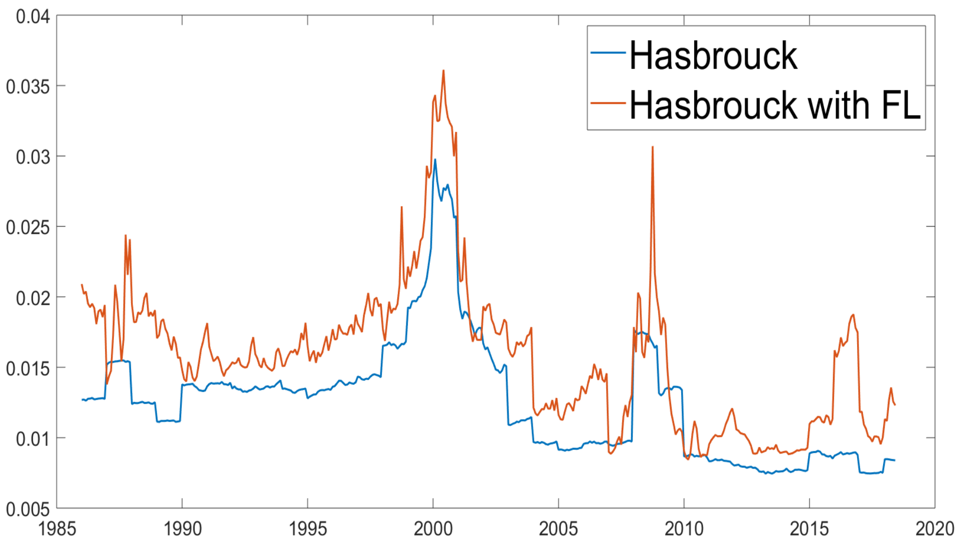


Figure 1.7: **Dynamics of transaction costs from H-Model and FLH-Model for high-volatility stocks**

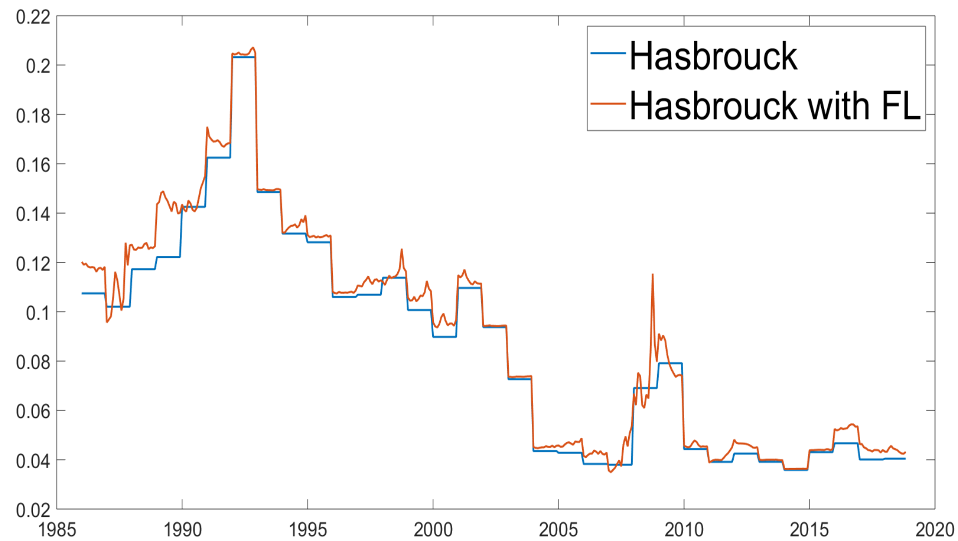


Figure 1.8: **Dynamics of transaction costs from H-Model and FLH-Model for low-volatility stocks**

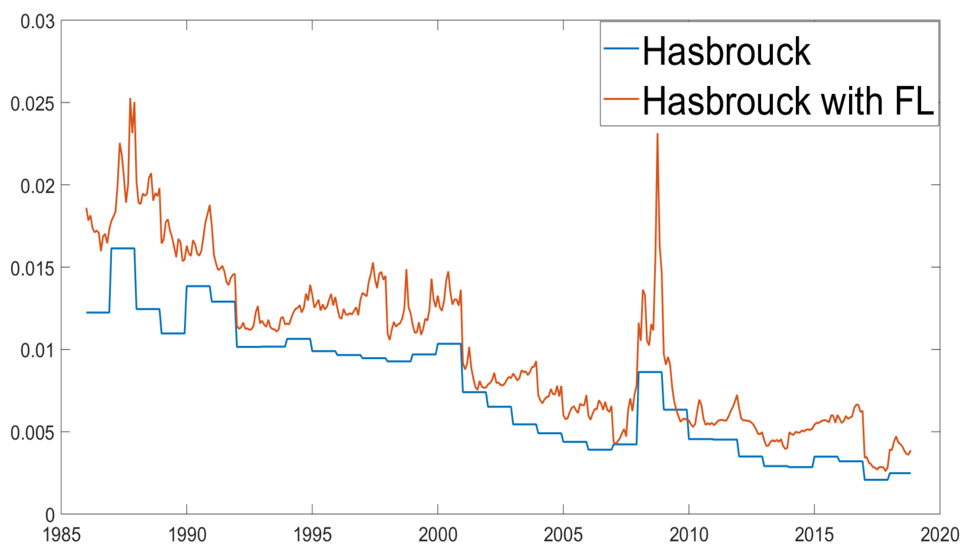


Figure 1.9: **Separating the transaction cost into its fixed component and its time-varying TED-spread component**

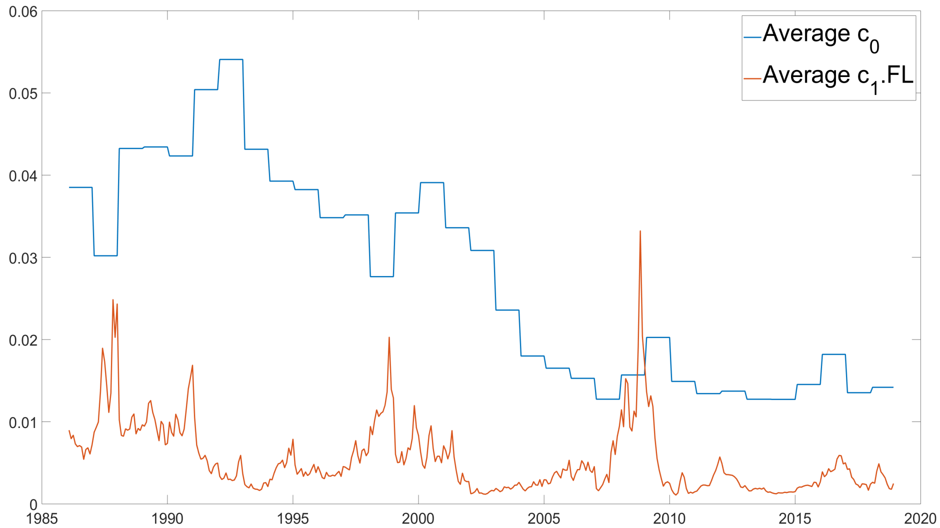


Figure 1.10: **Dynamics of transaction costs from H-Model and VIXH-Model for small, big, low-volatility, and high-volatility firms**

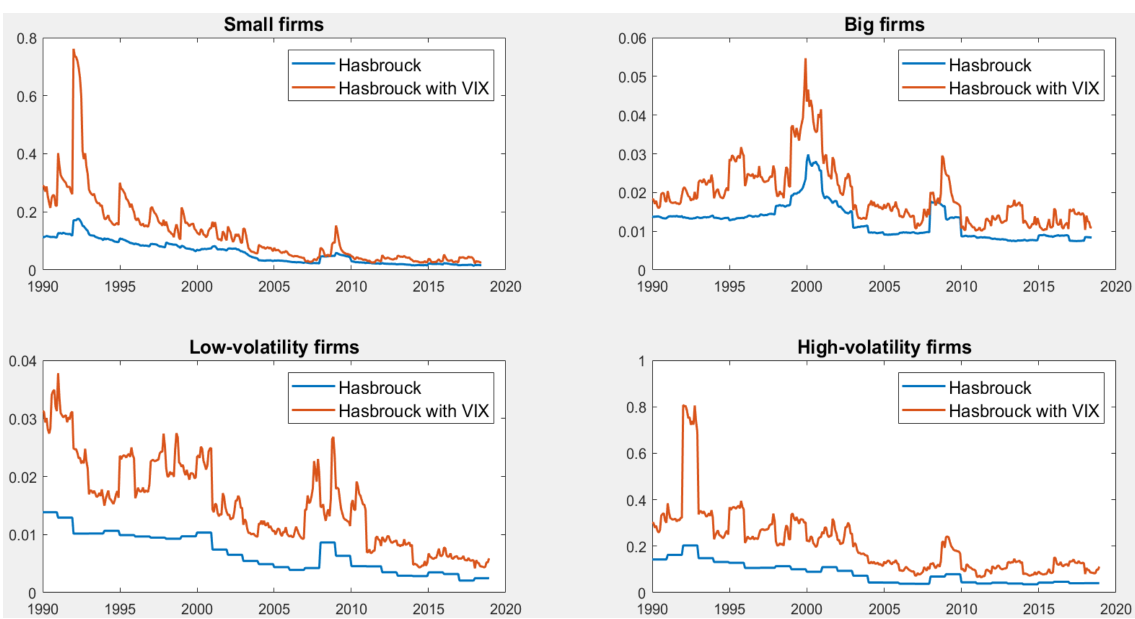
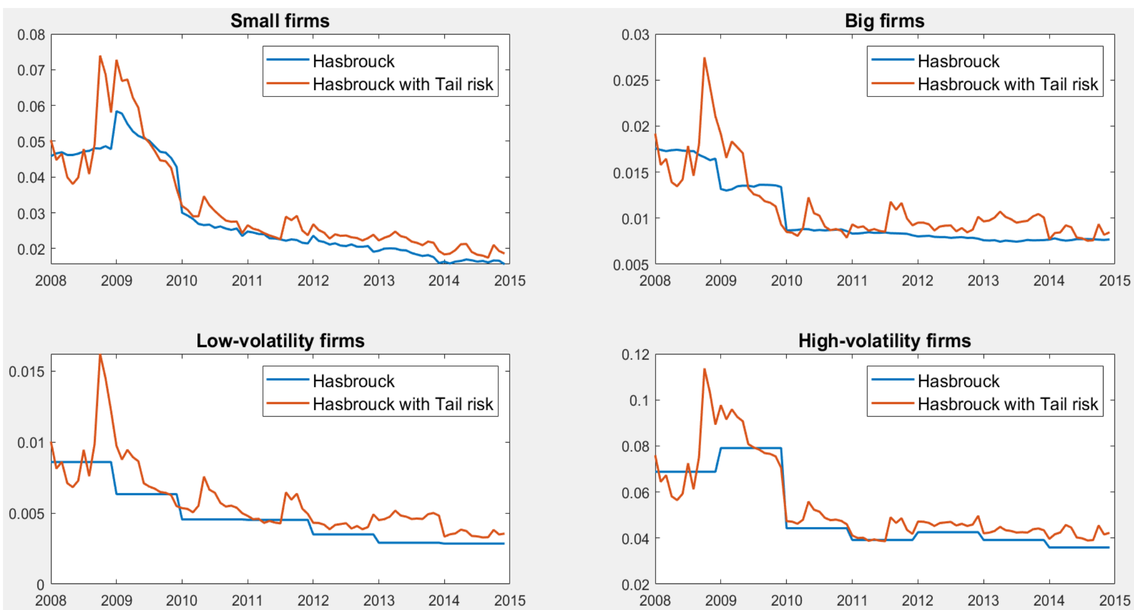


Figure 1.11: **Dynamics of transaction costs from H-Model and TRH-Model for small, big, low-volatility, and high-volatility firms**



Appendices for Chapter 1 (A)

A1 - Anomalies

Our anomaly definitions and descriptions are based on the lists of characteristics compiled by Novy-Marx and Velikov [2016] and Kozak et al. [2019].

- **Size (SIZE)**:

  Follows Fama and French [1993]. $SIZE = ME_{Jun}$. We use the CRSP end of June price times shares outstanding.

- **Realized Volatility (REALVOL)**:

  Follows Ang et al. [2006b]. $REALVOL = \frac{252}{N}\sum_{t=1}^{N} r_t^2$. N is the number of available returns for the stock for the given year. Rebalanced annually.

- **Industry Momentum (INDMOM)**:

  Follows Moskowitz and Grinblatt [1999b]. $INDMOM = rank(\sum_{l=1}^{6} r_{t-l}^{ind})$. In each month, the Fama and French 49 industries are ranked on their value-weighted past 6-months performance. Rebalanced monthly.

- **Share Issuance (annual) (NISSA)**:

  Follows Pontiff and Woodgate [2008]. $NISSA = shrout_{Jun}/shrout_{Jun-12}$, where $shrout$ is the number of shares outstanding. Change in real number of shares outstanding from past June to June of the previous year. Excludes changes in shares due to stock dividends and splits, and companies with no changes in $shrout$.

- **Composite Issuance (CISS)**:

Follows Daniel and Titman [2006]. $CISS = \log(\frac{ME_{t-13}}{ME_{t-60}}) - \sum_{l=13}^{60} r_{t-l}$, where r is the log return on the stock and ME is total market equity. Updated monthly.

- **Momentum (6m) (MOM6)**:

  Follows Jegadeesh and Titman [1993a]. $MOM6 = \sum_{l=2}^{7} r_{t-l}$. Cumulated past performance in the previous 6 months by skipping the most recent month. Rebalanced monthly.

- **Momentum (11m) (MOM11)**:

  Follows Jegadeesh and Titman [1993a]. $MOM11 = \sum_{l=2}^{12} r_{t-l}$. Cumulated past performance in the previous 11 months by skipping the most recent month. Rebalanced monthly.

- **Long-term Reversals (LTREV)**:

  Follows De Bondt and Thaler [1985]. $LTREV = \sum_{l=13}^{60} r_{t-l}$. Cumulative returns from $t-60$ to $t-13$. Updated monthly.

- **Short-term Reversal (STREV)**:

  Follows Jegadeesh [1990]. $STREV = r_{t-1}$. Return in the previous month. Updated monthly.

- **Seasonality (SEASON)**:

  Follows Heston and Sadka [2008]. $SEASON = \sum_{l=1}^{5} r_{t-l\times 12}$. Average monthly return in the same calendar month over the last 5 years. As an example, the average return from prior Octobers is used to predict returns this October. The firm needs at least one year of data to be included in the sample. Updated monthly.

- **Momentum-Reversal (MOMREV)**:

  Follows Jegadeesh and Titman [1993a]. $MOMREV = \sum_{l=14}^{19} r_{t-l}$. Buy and hold returns from $t-19$ to $t-14$. Updated monthly.

- **Share Issuance (monthly) (NISSM)**:

  Follows Pontiff and Woodgate [2008]. $NISSM = shrout_{t-13}/shrout_{t-1}$, where $shrout$ is the number of shares outstanding. Change in real number of shares outstanding from $t-13$ to $t-1$. Excludes changes in shares due to stock dividends and splits, and companies with no changes in $shrout$.

- **Industry Relative Reversals (INDRREV)**:

  Follows Da et al. [2013]. $INDRREV = r_{-1} - r_{-1}^{ind}$, where r is the return on a stock and r^{ind} is return on its industry. Difference between a stocks'prior month's return and the prior month's return of its industry (based on the Fama and French 49 industries). Updated monthly.

- **Price (PRICE)**:

  Follows Blume and Husic [1973]. $PRICE = log(ME/shrout)$, where ME is market equity and $shrout$ is the number of shares outstanding. Log of stock price. Updated monthly.

- **Share Volume (SHVOL)**:

  Follows Datar et al. [1998]. $SHVOL = \frac{1}{3}\sum_{i=1}^{3} volume_{t-i}/shrout_t$. Average number of shares traded over the previous three months scaled by shares outstanding. Updated monthly.

- **Investment-to-Capital (IK)**:

  Follows Xing [2007]. $IG = CAPX/PPENT$. Investment to capital is the ratio of capital expenditure ($CAPX$) over property, plant, and equipment ($PPENT$).

- **Investment Growth (IG)**:

  Follows Xing [2007]. $IG = CAPX/CAPX_{-12}$. Investment growth is the percentage change in capital expenditure (Compustat item CAPX).

- **Net Operating Assets (NOA)**:

  Follows Hirshleifer et al. [2004].

  $$NOA = (AT - CHE) - (AT - DLC - DLTT - MIB - PSTK - CEQ),$$

  where AT is total assets, CHE is cash and short-term investments, DLC is debt in current liabilities, $DLTT$ is long term debt, MIB is non-controlling interest, $PSTK$ is preferred capital stock, and CEQ is common equity. Updated annually.

- **Asset Growth (AG)**:

  Follows Cooper et al. [2008]. $AG = AT/AT_{-12}$. Rebalanced annually.

- **Investment-to-Assets (IA)**:

  Follows Chen et al. [2011]. $IA = \frac{PPEGT - PPEGT_{-12} + INVT - INVT_{-12}}{ATQ_{-12}}$. Investment-to-Assets is the annual change in $PPEGTQ$ which is property, plant, and equipment (Compustat item PPEGT) plus annual change in $INVT$ which is total inventories (Compustat item INVT) divided by lagged total assets (AT).

- **Leverage (LEV)**:

  Follows Bhandari [1988]. $LEV = (AT/ME)_{Dec}$. Market leverage is the ratio of total assets (Compustat item AT) over the market value of equity. Both are measured in December of the same year.

- **Return on Assets (annual) (ROAA)**:

  Follows Chen et al. [2011]. $ROAA = IB/AT$. Net income scaled by total assets. Updated annually.

- **Standardized Unexpected Earnings (SUE)**:

  Follows Foster et al. [1984]. $SUE = \frac{IBQ - IBQ_{-12}}{\sigma_{IBQ_{-24}:IBQ_{-3}}}$, where IBQ is income before extraordinary items (updated quarterly), and $\sigma_{IBQ_{-24}:IBQ_{-3}}$ is the standard deviation of IBQ in the past two years skipping the most recent quarter. Earnings surprises are measured by Standardized Unexpected Earnings (SUE), which is the change in the most recently announced quarterly earnings per share from its value announced four quarters ago divided by the standard deviation of this change in quarterly earnings over the prior eight quarters. Rebalanced monthly.

- **Return on Market Equity (ROME)**:

  Follows Chen et al. [2011]. $ROME = IBQ/ME_{-4}$, where IBQ is income before extraordinary items (updated quarterly), and ME is market value of equity. Rebalanced monthly.

- **Return on Book Equity (ROBE)**:

  Follows Chen et al. [2011]. $ROBE = IBQ/BEQ_{-3}$, where IBQ is income before extraordinary items (updated quarterly), and BEQ is book value of equity. Rebalanced monthly.

- **Sales Growth (SG)**:

Follows Lakonishok et al. [1994]. $SG = SALE/SALE_{-12}$. Sales growth is the percent change in net sales over turnover (Compustat item SALE).

- **Gross Profitability (GPROF)**:

  Follows Novy-Marx [2013]. $GPROF = GP/AT$, where GP is gross profits and AT is total assets. Rebalanced annually.

- **Gross Margins (GMARGINS)**:

  Follows Novy-Marx [2013]. $GMARGINS = GP/SALE$, where GP is gross profits and $SALE$ is total revenues. Rebalanced annually.

- **Piotroski's F-score (FSCORE)**:

  Follows Piotroski [2001]. $FSCORE = 1_{IB>0} + 1_{\Delta ROA>0} + 1_{CFO>0} + 1_{CFO>IB} + 1_{\Delta DTA<0|DLTT=0|DLTT_{-12}=}$ $1_{\Delta ATL>0} + 1_{EqIss\leq 0} + 1_{\Delta GM>0} + 1_{\Delta ATO>0}$, where IB is income before extraordinary items, ROA is income before extraordinary items scaled by lagged total assets, CFO is cash flow from operations, DTA is total long-term debt scaled by total assets, $DLTT$ is total long-term debt, ATL is total current assets scaled by total current liabilities, $EqIss$ is the difference between sales of common stock and purchases of common stock recorded on the cash flow statement, GM equals one minus the ratio of cost of goods sold and total revenues, and ATO equals total revenues, scaled by total assets. Rebalanced annualy.

- **Asset Turnover (ATURNOVER)**:

  Follows Soliman [2008]. $ATURNOVER = SALE/AT$. Sales to total assets. Rebalanced annually.

- **Sales-to-Price (SP)**:

  Follows Barbee Jr et al. [1996]. $SP = SALE/ME_{Dec}$. Total revenues divided by stock price. Updated annually.

- **Accruals (ACC)**:

  Follows Sloan [1996].

$$ACC = \frac{\Delta ACT - \Delta CHE - \Delta LCT + \Delta DLC + \Delta TXP - \Delta DP}{(AT + AT_{-12})/2}$$

, where ΔACT is the annual change in total current assets, ΔCHE is the annual change in total cash and short-term investments, ΔLCT is the annual change in current liabilities, ΔDLC is the annual change in debt in current liabilities, ΔTXP is the annual change in income taxes payable, ΔDP is the annual change in depreciation and amortization, and $(AT + AT_{-12})/2$ is average total assets over the last two years. Rebalanced annually.

- **Growth in Long Term Net Operating Assets (GLTNOA)**:

  Follows Fairfield et al. [2003]. $GLTNOA = GRNOA - ACC$. Growth in Net Operating Assets minus Accruals. $NOA = (RECT + INVT + ACO + PPENT + INTAN + AO - AP - LCO - LO)/AT$, $GRNOA = NOA - NOA_{-12}$, $ACC = ((RECT - RECT_{-12}) + (INVT - INVT_{-12}) + (ACO - ACO_{-12}) - (AP - AP_{-12}) - (LCO - LCO_{-12}) - DP)/((AT + AT_{-12})/2)$, where $RECT$ = Receivables, $INVT$ = Total Inventory, ACO = Current Assets, AP =Accounts Payable, LCO = Current Liabilities (Other), DP = Depreciation and Amortization, AT = Assets, $PPENT$ = Property, Plant, and Equipment (net), $INTAN$ = Intangible Assets, AO = Assets (Other), LO = Liabilities (Other). Updated annually.

A2 - Additional Tables and Figures

Table 1.11: **VIXH-Model: Transaction costs and flight to quality**

The table features the average absolute change and the average change in percentage between the transaction costs estimated with the Hasbrouck [2009] model with the VIX (VIXH-Model) and the Hasbrouck model (H-Model). For the two anomalies, size and volatility, we report the changes for the ten decile portfolios. We also compute the average parameter c_1 per portfolio since this parameter measures the sensitivity of the portfolio transaction cost to the VIX. We perform an ANOVA to test the difference of all these values across the deciles.

	Size			Volatility		
	Absolute change	Change in percentage	Parameter c_1	Absolute change	Change in percentage	Parameter c_1
Dec1	0.0651	91.84	1.5430	0.0084	129.33	0.1825
Dec2	0.0304	70.09	0.8424	0.0083	90.12	0.2240
Dec3	0.0226	68.35	0.6501	0.0093	76.65	0.2683
Dec4	0.0173	64.94	0.5243	0.0117	73.76	0.3337
Dec5	0.0147	65.01	0.4547	0.0156	82.59	0.4209
Dec6	0.0133	65.55	0.4076	0.0174	75.02	0.4875
Dec7	0.0118	65.24	0.3666	0.0214	77.57	0.5919
Dec8	0.0103	63.12	0.3290	0.0277	78.36	0.7539
Dec9	0.0087	60.82	0.2854	0.0409	86.45	1.0550
Dec10	0.0067	54.85	0.2466	0.1257	151.66	2.7129
Number of periods	342 months	342 months	29 years	348 months	348 months	29 years
Anova: F stat	113.26***	24.60***	12.80***	370.54***	148.29***	31.86***
Anova: DF Columns	9	9	9	9	9	9
Anova: DF Errors	3410	3410	280	3470	3470	280

Table 1.12: **TRH-Model: Transaction costs and flight to quality**

The table features the average absolute change and the average change in percentage between the transaction costs estimated with the Hasbrouck [2009] model with tail risk (TRH-Model) and the Hasbrouck model (H-Model). For the two anomalies, size and volatility, we report the changes for the ten decile portfolios. We also compute the average parameter c_1 per portfolio since this parameter measures the sensitivity of the portfolio transaction cost to the tail risk measure of Weller [2019]. We perform an ANOVA to test the difference of all these values across the deciles.

	Size			Volatility		
	Absolute change	Change in percentage	Parameter c_1	Absolute change	Change in percentage	Parameter c_1
Dec1	0.0029	11.65	358.74	0.0010	25.34	104.30
Dec2	0.0023	10.31	334.45	0.0011	21.19	139.12
Dec3	0.0021	11.65	292.65	0.0012	17.53	166.57
Dec4	0.0018	11.75	286.11	0.0013	14.69	198.89
Dec5	0.0016	11.67	294.34	0.0015	13.79	236.96
Dec6	0.0015	11.95	278.13	0.0017	12.92	271.35
Dec7	0.0014	11.61	260.07	0.0019	11.84	316.61
Dec8	0.0013	12.10	245.49	0.0023	11.93	368.78
Dec9	0.0013	13.45	219.43	0.0030	11.51	445.70
Dec10	0.0011	12.56	194.96	0.0046	9.81	555.46
Number of periods	84 months	84 months	7 years	84 months	84 months	7 years
Anova: F stat	2.20**	0.20	2.59**	6.77***	6.26***	18.62***
Anova: DF Columns	9	9	9	9	9	9
Anova: DF Errors	830	830	60	830	830	60

Figure 1.12: **Dynamics of transaction costs from H-Model and FLH-Model for Losers**

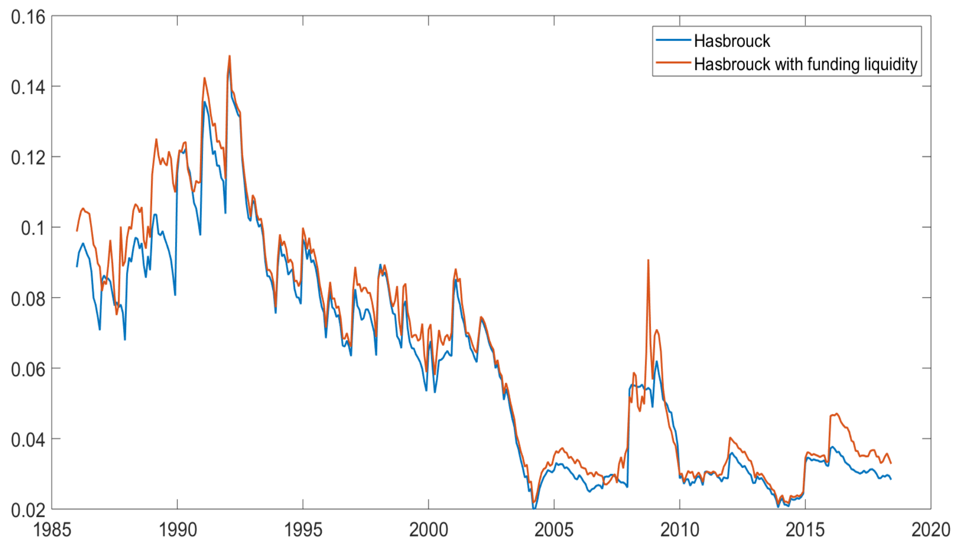


Figure 1.13: **Dynamics of transaction costs from H-Model and FLH-Model for Winners**

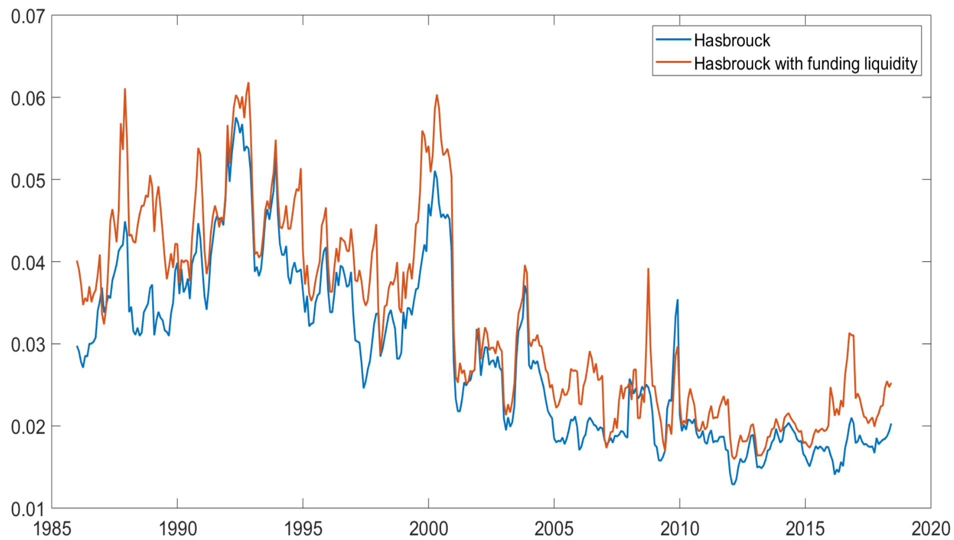


Figure 1.14: **Dynamics of transaction costs from H-Model and VIXH-Model for COCA COLA CO**

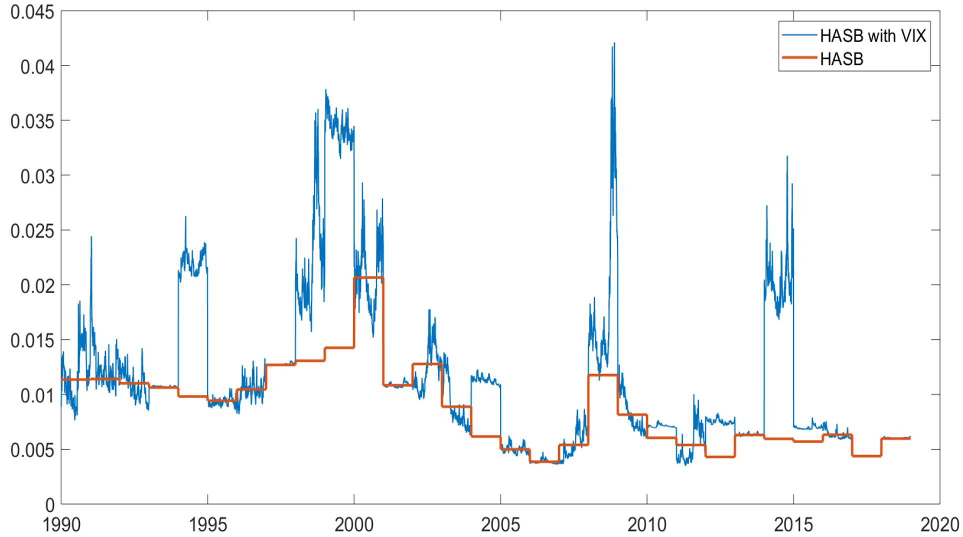


Figure 1.15: **Dynamics of transaction costs from H-Model and VIXH-Model for ROCKY MOUNTAIN CHOCOLATE FACTORY**

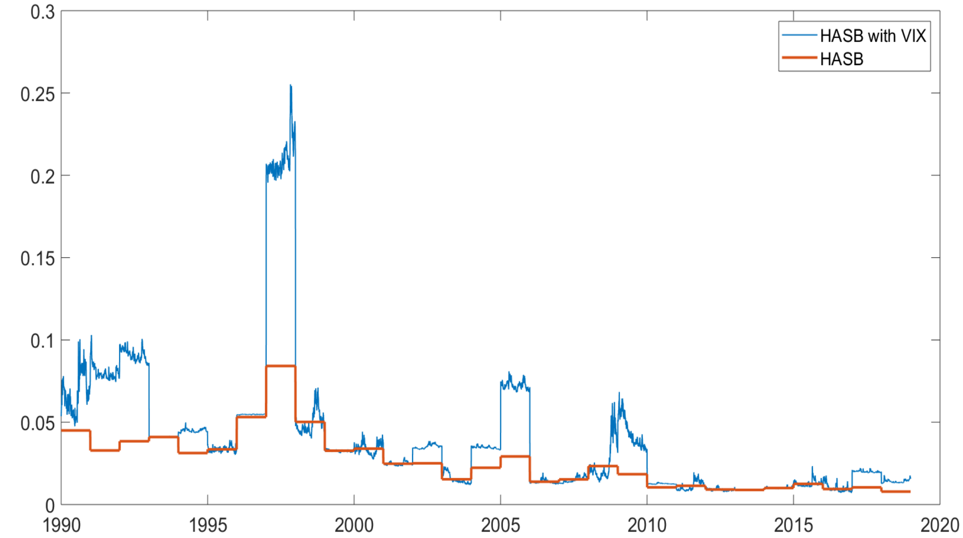


Figure 1.16: **Dynamics of transaction costs from H-Model and TRH-Model for COCA COLA CO**

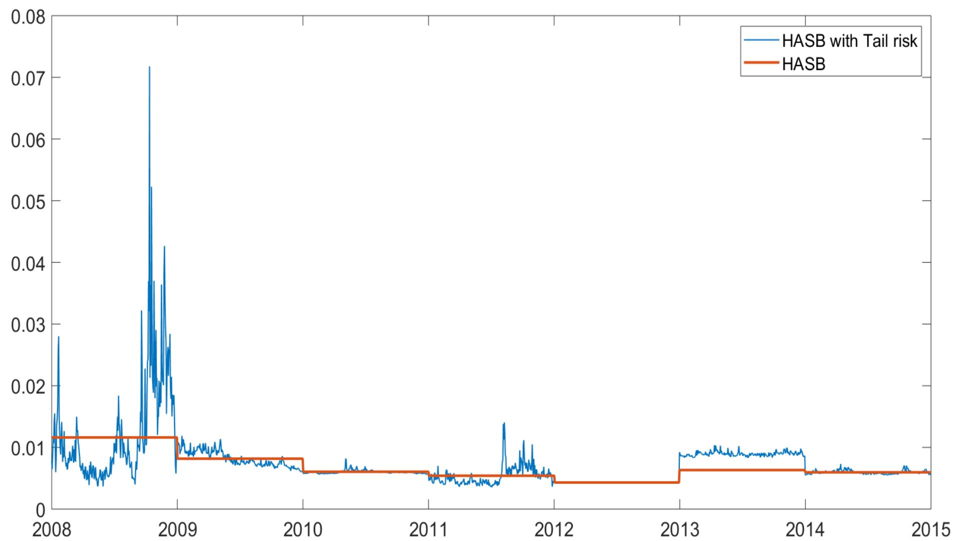


Figure 1.17: **Dynamics of transaction costs from H-Model and TRH-Model for ROCKY MOUNTAIN CHOCOLATE FACTORY**

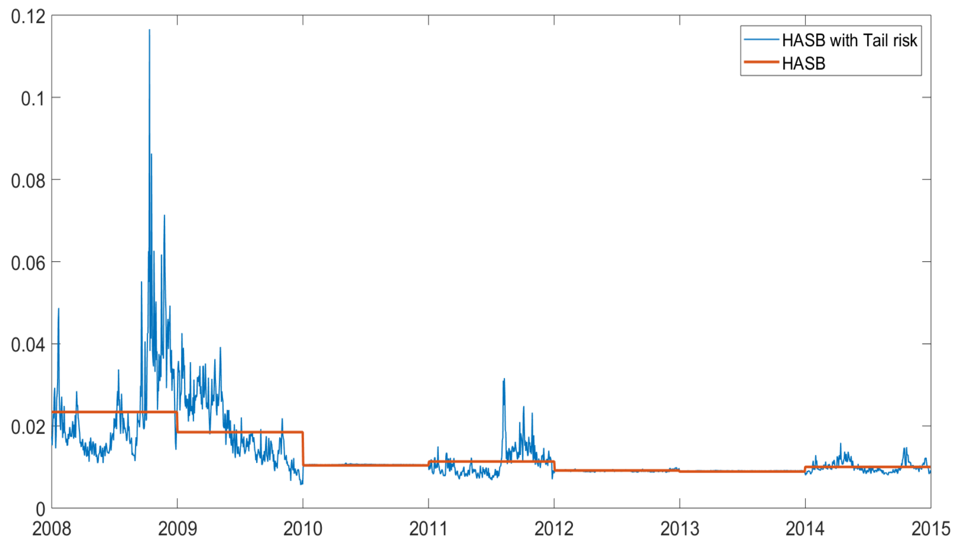


Figure 1.18: **Separating the transaction cost into its fixed component and its time-varying TED-spread component: small firms**

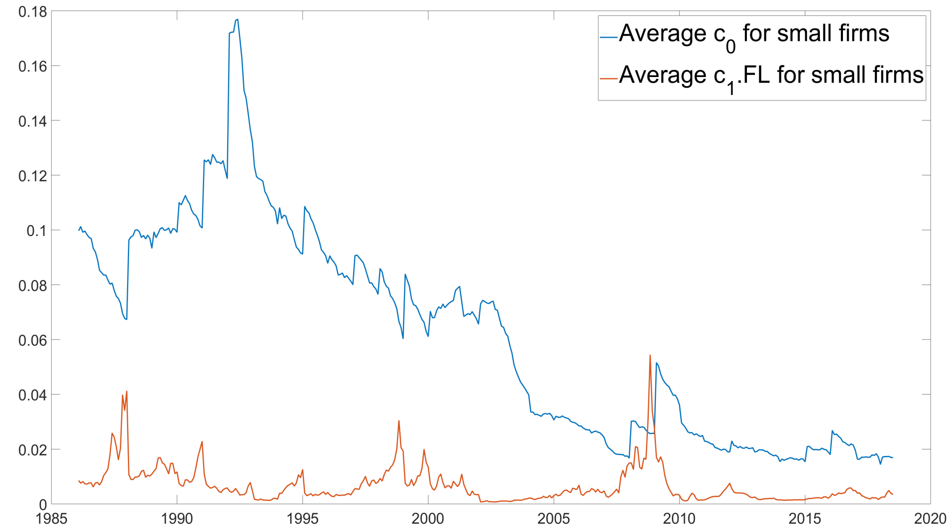
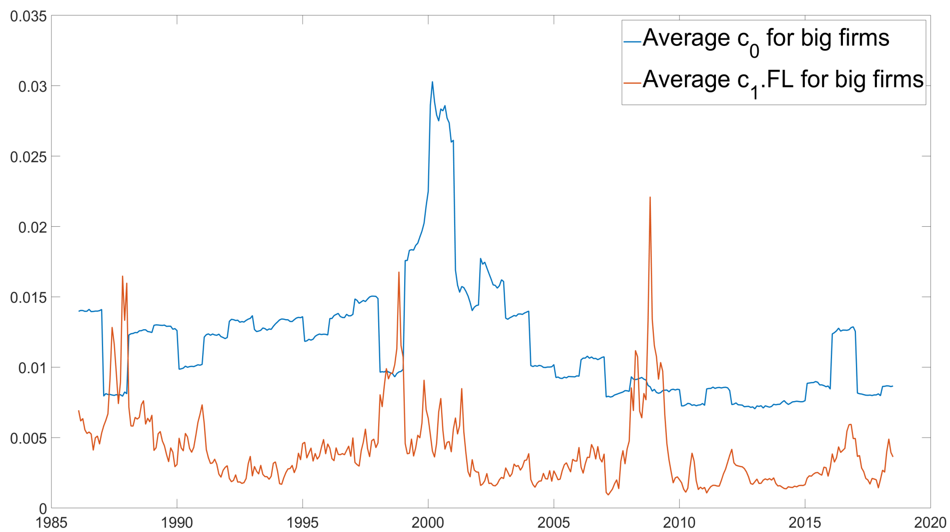


Figure 1.19: **Separating the transaction cost into its fixed component and its time-varying TED-spread component: large firms**



Chapter 2

Journal Picking for Better Returns

2.1 Introduction

Many studies have documented the importance of academic research in the finance field. In fact, investors base their strategies on academic publications that focus on anomalies. Anomalies are firm characteristics or other observable variables that provide explanatory power for the cross-section of sample mean returns beyond the beta of the CAPM or a benchmark factor model. To take advantage of anomalies, arbitrageurs pursue long-short strategies. They buy stocks that are the most exposed to the anomaly (for example small firms) and sell the ones that are the least exposed (large firms) to capture their difference in returns. This paper studies how popular anomalies are in peer-reviewed finance journals and how this popularity can influence the returns of strategies based on these anomalies. More precisely, this paper uses the tone of publications in which anomalies appear and the impact factor of journals in which the publications appear to forecast the returns of strategies based on anomalies.

These two objectives are motivated by recent contributions to the literature on the importance of academic research on anomalies for investors. Anomalies could be either the result of cross-sectional differences in risk (Fama [1991], Fama [1998]), mispricing (Barberis and Thaler [2003]) or data mining (Fama [1998]). Engelberg et al. [2018] show that anomalies are more likely the results of mispricing. McLean and Pontiff [2016] study the post-publication return of 97 anomalies and find portfolio returns are 58% lower post-publication, suggesting investors learn about mispricing from academic publications. Academic publications therefore contribute to the destruction of returns attached to anomalies because once investors learn about an anomaly, they will trade this anomaly and arbitrage away the mispricing. We can make an analogy with a diamond mine. Once people learn about the existence of this mine, it is very likely that the diamond reserves of the mine will decrease in the coming periods.

This paper differs from this recent literature both in its methodology and its scope. In their study, McLean and Pontiff [2016] did not study how popular an anomaly becomes in the literature after its first publication and how this popularity affects strategies based on this anomaly. The idea here is the fact that, once a new anomaly is published, there is a rush of investors towards strategies based on this anomaly, but things can change as time goes by. In fact, some other studies either by using different sample periods, different econometric models or by taking in account transaction costs, find that some anomalies are actually not profitable (Kim and Kim [2003], Choi et al. [2016], Lesmond et al. [2004], Liu et al. [2018], Zhu [2012]...). Positive feedback from publications should therefore destroy anomalies returns by attracting more investors that will arbitrage away the mispricing. The opposite effect should be observed with negative feedback.

To achieve its research objectives, the first step of this paper consists of using textual data to construct the tone index for each anomaly and for each month. It uses bibliometric data on articles published in 23 major finance, accounting, econometric, and economic journals. These data come from the Web of Science website. For each of the 23 anomalies that this paper takes into account, I identify the publications that discuss the anomaly. I then use the dictionary of positive and negative words proposed by Loughran and McDonald [2011] to get the tone of the abstract of the publication. The tone in this paper is measured either by the difference between the number of positive words and the number of negative words in the abstract, or by the difference between the percentage of positive words and the percentage of negative words in the abstract. I also take in account the impact factor of the journals in which the publications appear. The journal impact factor is the yearly average number of citations that articles published in the last two years in a given journal received. It is frequently used as a proxy for the relative importance of a journal within its field; journals with higher impact factors are often deemed to be more important than those with lower ones. The idea of using the impact factor here is the fact that conclusions from the publications that appear in a high impact factor journal are more likely to have an impact on investors' behaviour.

The second step consists of using the impact factor of the journal in which the publication that discusses the anomaly is published and the tone index to forecast returns of long-short anomaly-based portfolios. These portfolios are constructed for a set of anomalies considered one at a time. Each month, stocks are ranked based on the value of the anomaly. Stocks are then grouped into deciles. The long-short portfolio is then obtained by going long on the stocks in the highest decile and shorting the stocks in the lowest decile or inversely, depending on the anomaly[1]. For each anomaly a, I consider months $m_{(a,1)}, m_{(a,2)}, \ldots, m_{(a,n_a)}$, for which there exist at least a publication that discusses this anomaly a. n_a is the number of distinct months for which at least a publication discusses anomaly a. For each month $m_{(a,i)}, i = 1, 2, \ldots, n_a$, I compute $JIF_{a,m_{(a,i)}}$, the average impact factor of journals in which publications that discussed anomaly a appeared, and $TI_{a,m_{(a,i)}}$, the average tone index of abstracts of publications that discussed anomaly a. I then construct the average return of the long-short portfolio based on anomaly a for periods $m_{(a,1)}$ to $m_{(a,2)}$, ..., $m_{(a,n_a-1)}$ to $m_{(a,n_a)}$ and $m_{(a,n_a)}$ to 2018-06. I call these returns *Between publication months average returns.* 1985-01 to 2018-06 is the sample period in this paper. For each month $m_{(a,i)}$, $TI_{a,m_{(a,i)}}$ and $JIF_{a,m_{(a,i)}}$ are used to predict the average return of the long-short portfolio based on anomaly a over the period going from $m_{(a,i)}$ to $m_{(a,i+1)}$, or 2018-06 if $i = n_a$. In fact, if during a month, there is at least one publication that discusses anomaly a, the conclusions of those publications are going to affect the behaviour of investors until the next month anomaly a appears in a publication. Therefore, I suppose the behaviour of investors are going to change as long as new publications about an anomaly are going to appear. So, for anomaly a, the larger the spacing between the publication months, the more stable the investors' behavior and this will be reflected in the returns of strategies based on the anomaly.

Two regressions are run. The first one is a panel regression in which the dependent variable is the post-publication returns. For each anomaly, the return is regressed on the average tone index and the journal impact factor of the latest month in which the anomaly appears in a publication. In the second regression, the dependent variable is the *Between publication months average returns* and the covariates are the tone index and the journal impact factor. In the two regressions, I include an interaction term between the covariates because a positive or a negative conclusion of a publication (tone index) will probably have more effect if the publication appears in a journal with a high impact factor. I have run the

[1]It depends on the anomaly. For momentum for example, the portfolio is obtained by going long on the stocks in the highest decile and shorting those in the lowest decile while for the size, the portfolio is built obtained by going long on the stocks in the lowest decile and shorting those in the highest decile

second regression multiple times by getting rid of observations for which the interval between two publications months is less than a given number q of months $(m_{(a,i+1)} - m_{(a,i)})$. Results of the second regression are obtained for different values of q, $(q = 1, 2, 3, ...)$.

Empirical results show that the anomalies with the highest average tone index are the dividend yield, the long term reversal and the investment to capital, while the anomalies with the lowest average tone index are the low-volatility anomaly, the sales growth and the asset growth. Most of them are published in top finance journals like Journal of Finance, Journal of Financial Economics and Review of Financial studies.

Empirical results of the regressions show that the tone index has a significant and positive coefficient on the post publication returns and on the *Between publication months average returns*, the coefficient of the journal impact factor is positive and the coefficient of the interaction is negative. The overall fitness of the second regression measured by the adjusted R^2 gets better when we focus more on observations for which the interval between two publications months is larger, i.e. q is larger. This proves that the greater the spacing between the publication months, the more stable the investors' behavior and this is reflected in the returns of strategies based on the anomaly. The idea that academic publications destroy anomalies returns discussed in McLean and Pontiff [2016] is observed here when publications about anomalies happen is high impact factor journals. In fact, when an anomaly is discussed in a positive tone publication that appears in a journal with an impact factor higher than 3 (Journal of Finance, Journal of Financial Economics, Review of Financial Studies), this anomaly is more likely to attract investors that are going to arbitrage away the mispricing and therefore destroy the anomaly.

The rest of the paper is structured as follows. In Section 2.2, I describe the methodology of research. In this section,I explained how I identify publications that I need according to anomalies,and how I construct the index. I also describe how I use the index of popularity to forecast the returns on long-short anomaly-based portfolios. Section 2.3 describes all the data sources needed in this research. Section 2.4 presents the results of the index of popularity and the results of the forecasting of the anomaly returns with the index of popularity. In Section 2.5 I do some robustness checks by considering only publications that appear in Journal of Banking and Finance, and also by accounting for the transaction costs. Section 2.6 concludes.

Related Literature

This paper is related to many papers in the literature. Many papers have shown the effect of finance publications on anomalies. McLean and Pontiff [2016] study the post-publication return predictability of 97 variables shown to predict cross-sectional stock returns. Their study finds that portfolio returns are 58% lower post-publication. Jacobs and Müller [2020] study the pre- and post-publication return predictability of 241 cross-sectional anomalies in 39 stock markets. They find, based on more than two million anomaly country-months, that the United States is the only country with a reliable post-publication decline in long-short returns. This paper differs from those two papers as it did not just focus on the first time anomalies get published in academic journals, but tracks every month there is a publication about the anomaly.

This paper is also related to the large set of papers in finance that use textual data to predict returns (Loughran and McDonald [2011], Hillert et al. [2014], Heston and Sinha [2017]...). Usually papers in finance that use textual analysis take textual data from annual reports / 10-Ks / 10-Qs (Loughran and McDonald [2011]), from Earnings press Releases / Earnings conference calls (Henry [2008]), from media[2] (Garcia [2013], Tetlock [2007], Tetlock et al. [2008]) or from internet expressed sentiment(Das and Chen [2007]). This paper differs from all those papers as it is the first to use articles from academic peer-reviewed journals to forecast returns.

2.2 Methodology

2.2.1 Publications that discuss an Anomaly

The first thing to do in this study is to identify publications that discuss one of the 23 anomalies taken into account. Table 1.4 presents these 23 anomalies. To do so, this paper uses the title, the keywords and the abstract of the publication. It focuses on publications that not only mention the anomaly in their abstract but also feature some cross-sectional studies between the anomaly and future stock returns. To check if an anomaly is discussed in a given paper, I use textual analysis and check if words related to the given anomaly appear in the title or the abstract of the publication. I did not use the body of the publication

[2]News stories and commentaries, Analyst reports

because words related to an anomaly could appear in the body of the publication and still the publication is not discussing the given anomaly. While, for the title and the abstract, words are chosen wisely to summarize what is being discussed in the publication. So, if words related to a particular anomaly appear in the title or the abstract of a publication, there is a strong chance the publication is discussing this particular anomaly.

A publication p_m that appears in month m, is said to discuss an anomaly a ($a \in p_m$) if it has in its title or its abstract the name of the anomaly and has in its title, keywords or abstract one of the following words: anomaly, abnormal return, long-short portfolio returns, fundamental analysis, return predictability, cross-section, strategies, security return, stock return, portfolio, premium. While searching the name of the anomaly through the title, the keywords and the abstract of the publication, this paper takes into account the fact that the anomaly can be denominated differently across publications. For example, "standard unexpected earning", "postearnings-announcement drift", "earnings surprise" ... all refer to the same anomaly. The same goes for "residual variance" and "idiosyncratic volatility".

2.2.2 The Tone Index and the Journal Impact Factor

The information needed from the publications that discuss anomalies are the tone of the publication and also the impact factor of the journal in which the publication appears. For a given anomaly a and a publication p, in order to get the information needed, three questions need to be answered:

- Is the publication talking about the anomaly?

  This question is necessary because the more an anomaly is discussed in publications, the more the anomaly will tend to be popular. For an investor to be interested in an anomaly and design a strategy based on this anomaly, she needs first of all to hear about the anomaly.

- What is the tone of the publication?

  This second question is also necessary, because an anomaly could be discussed in a publication but still, conclusions about this anomaly will be negative. Let us consider the paper of Lesmond et al. [2004] for example. In this paper, the authors show that the magnitude of the abnormal returns associated with momentum strategies creates an illusion of profit opportunity when, in fact, none exists when transaction costs are taken into account. After reading such paper, an investor will probably decide not to use strategies based on momentum.

- What is the impact factor of the journal in which the publication appears?

  The impact factor is important because a publication that appears in a journal with a high impact factor has more chances to be known about than a publication that appears in a journal with a lower impact factor. Plus, Investors will probably trust results from a journal with a high impact factor.

After checking if the publication is discussing the anomaly, I now need to get the tone of the publication. For this purpose, the sentiment analysis is used. Sentiment analysis is a technique used in textual analysis to determine the tone (positive, neutral or negative) of a text. Results from many papers in finance such as Antweiler and Frank [2004], Tetlock [2007], Engelberg [2008], Li [2008] and Tetlock et al. [2008] indicate that negative word classications can be effective in measuring tone, as reflected by signicant correlations with other financial variables.

The sentiment of a text is determined by the sentiment of the words in the text. While determining the sentiment of words may be as difficult as determining the text's sentiment, there exist dictionaries built by others that associate words to their sentiments. A commonly used source for word classications is the Harvard Psychosociological Dictionary, specically, the Harvard-IV-4 TagNeg (H4N) le. Loughran and McDonald [2011] show that this dictionary is not suitable for specific topics like finance and propose a new dictionary that better reflects tone in financial texts. On top of that, unlike the Harvard Psychosociological Dictionary, the dictionary proposed by Loughran and McDonald [2011] account for inflections, or different forms of the same word. For example, if we consider *aberrant* a negative word, we would probably also want to include words such as *aberration*, *aberrational* and *aberrations* into the dictionary of negative words.

I have done some modifications to the dictionary proposed by Loughran and McDonald [2011] to fit this particular study. For example, in Loughran and McDonald [2011] dictionary of negative words, there are words such as *abnormal*, *anomalies*, or *mispricing* that do not necessarily mean something negative for an investor. For an investor abnormal returns or mispriced assets are opportunities of making profits.

By using this dictionary, this paper computes the tone index as follows:

$$TI_{a,m} = \begin{cases} \sum_{p_m, a \in p_m} \frac{TI_{p_m}}{|\{p_m, a \in p_m\}|}, & \text{if } |\{p_m, a \in p_m\}| > 0, \\ 0 & \text{if } |\{p_m, a \in p_m\}| = 0, \end{cases} \tag{2.1}$$

where, TI_{p_m}, the tone index is either the difference of the number of positive words and the number of negative words in the abstract of publication p_m or the difference of the percentage of positive words and the percentage of negative words in the abstract of pubication p_m.

The journal impact factor of a journal j in year n is computed by Clarivate Analytics and is given by the following formula:

$$JIF_{j,n} = \frac{\text{Citations in year } n \text{ to items published in years } n-1 \text{ and } n-2 \text{ in journal } j}{\text{Number of citable items published in years } n-1 \text{ and } n-2 \text{ in journal } j} \tag{2.2}$$

2.2.3 Empirical Analyses

There are two main regression models in this paper. In the first regression, I have a panel regression in which the anomaly returns are regressed on the tone index and on the journal impact factor.

$$R_{a,t} = \alpha + \beta_1.TI_{p_{a,t}} + \beta_2.JIF_{a,t} + \beta_3.TI_{p_{a,t}}.JIF_{a,t} + \varepsilon_{a,t}. \tag{2.3}$$

In Equation 2.3, $R_{a,t}$ is the return of long-short strategy based on anomaly a at month t. $TI_{p_{a,t}}$ is the average tone index of the latest publications in which anomaly a has been discussed and $JIF_{a,t}$ is the average impact factor of the journals in which the latest publications that discuss anomaly a appear. By *latest publications* here, this paper means the publications that appear during the latest of the months t' previous to t, $t' < t$. $\varepsilon_{a,t}$ is the error term.

In the second regression equation, what is explained is the average return of the period going from the next month following a month in which at least a publication discussed an anomaly to the next month a publication that discusses the same anomaly appears.

$$R_{a,[m_i,m_{i+1}]} = \alpha + \beta_1.TI_{p_{a,m_i}} + \beta_2.JIF_{a,m_i} + \beta_3.TI_{p_{a,m_i}}.JIF_{a,m_i} + \varepsilon_{a,m_i}. \tag{2.4}$$

In Equation 2.4, $[m_i + 1, m_{i+1}]$ represents the interval between two successive months of appearance of anomaly a in a publication. TI_{a,m_i}, the average tone index of abstracts of publications that discussed anomaly a and JIF_{a,m_i} is the average impact factor of journals in which publications that discussed anomaly a appeared. $R_{a,[m_i+1,m_{i+1}]}$ is the *between publication months average return*, it is the average return of a long-short strategy based on anomaly a over the period $[m_i + 1, m_{i+1}]$.

I decide to use both the *between publication months average return* as dependent variable because publications about anomaly a are dynamic and information that comes from these publication will probably affect investors' behaviour until the next month in which there will be a publication that will discuss the same anomaly. If during a month, there is at least one publication that discusses anomaly a, the conclusions of those publications will be the current information about the anomaly and are going to affect the behaviour of investors until the next month anomaly a appears in a publication. Therefore, the behaviour of investors is going to change as long as new publications about an anomaly are going to appear and investors will adapt their strategies accordingly. So, for anomaly a, the larger the spacing between the publication months, the more stable the investors' behavior and this will be reflected in the returns of strategies based on the anomaly.

In those two equations, I include interactions between the tone index and the journal impact factor because a positive or a negative conclusion of a publication (tone index) will probably have more effect if the publication appears in a journal with a high impact factor. As mentioned before, when a publication that discusses a given anomaly appears, the longer the wait until the next publication about the same anomaly, the better it is, because information from this publication will have enough time to spread among investors. Therefore, the regression in Equation 2.4 is run multiple times by getting rid of observations for which the interval between two publications months is less than a given number q of months ($m_{(a,i+1)} - m_{(a,i)}$). The regressions are run for different values of q, ($q = 1, 2, 3, ...$). I expect to get better results as q increases.

The interpretation of the coefficients from the two equations is quite the same. $\beta_1 + \beta_3.JIF$ measures the effect of the tone on the returns or on the *between publication months average return*. The tone will be said to destroy stock return predictability if $\beta_1 + \beta_3.JIF$ is negative, meaning it has a negative impact on the return. $\beta_2 + \beta_3.TI$ measures the effect of the impact factor of the journal in which the publication that discusses the anomaly appears. A negative $\beta_2 + \beta_3.TI$ means anomalies are destroyed when they are published in high impact factor journals.

2.3 Data

This paper uses different data sources. First of all, I collect data from Web of Science (WoS) database and Together. This site constitutes one of the largest depository of academic research in economics. A web crawling algorithm is used to collect all the informations about publications. These informations are then organized into a novel database with variables such as, the title, the keywords, the abstract, the journal name, the year and month of publication etc. I take all academic publications in leading peer-reviewed finance and accounting journals. I have downloaded from Web of Science 67,061 publications of 23 journals publishing in finance and accounting. I also take in account journals such as Quaterly Journal of Economics and Journal of Political Economy that are not specialized in finance but have some publications discussing finance topics.I get rid of publications for which one of the following variables is not available: publication date, title, keywords or abstract. After that I am left with 32,978 publications. Table 2.1 presents the number of publications for each journal. In the database obtained, *Management Science* and *Journal of Banking and Finance* are the journals with most publications.

This paper also requires the journal impact factor. The impact factor of each journal concerned is obtained on the Website of Clarivate Analytics. The impact factor is available from 1997 to 2018. 1985-01 to 2018-06 is the sample period in this paper. So the time periods for regressions in which the journal impact factor is considered are adapted accordingly.

Anomalies are obtained via various sources. Some are computed following Novy-Marx and Velikov [2016] and Kozak et al. [2019] using two data sources: some anomalies have been computed using COMPUSTAT Data (North America - Fundamentals Quaterly) and some using CRSP. Some are obtained directly from the Data Library of the Kenneth French website. The anomalies used in this paper are presented in Appendix 2.6. I take in account the value weighted portfolios and the equally weighted portfolios.

2.4 Results

2.4.1 Tone Index and Journal Impact Factor

Figure 2.1 shows the evolution of the number of publications that discuss anomalies. As we can see, as time goes, the number of publications that discuss anomalies is increasing. But this upward trend disappears when I consider the percentage of publications that discuss anomalies (See Figure 2.2). The upward trend is only there between 1995 and 2006 and then it became stable.

Figure 2.1, Figure 2.2, and Table 2.2 here.

Table 2.2 presents each anomaly with the average number of positive words, the average number of negative words, the percentage of positive words, the percentage of negative words over the abstracts of publications in which it appears, and the average tone index measured by the difference of the percentage of positive words and the percentage of negative words. The anomalies that appear the most in publications are the dividend yield (144 publications), the value (111 publications) and the standardized unexpected earning (109 publications). The anomalies with the highest average tone index are the dividend yield, the long term reversal and the investment to capital, while the anomalies with the lowest average tone index are the low-volatility anomaly, the sales growth and the asset growth.

Table 2.3 here.

Table 2.3 presents each anomaly with the average impact factor of the journal in which the anomaly appears. In average the anomalies appears in average, in the same kind of journal. Most of them are published in top finance journals like Journal of Finance, Journal of Financial Economics and Review of Financial studies. A journal like Journal of Banking & Finance which is not as high ranked like Journal of Finance or Journal of Financial Economics also has many publications that discussed anomalies (See Table 2.1).

2.4.2 Portfolio Returns Relative to End-of-Sample and First Publication Dates

In this subsection I run the same regression as McLean and Pontiff [2016] in which they formally study the returns of each anomaly relative to its sample-end and publication dates. The baseline regression model is described in the following equation:

$$R_{a,t} = \alpha_i + \beta_1.\text{Post Sample Dummy}_{a,t} + \beta_2.\text{Post Publication Dummy}_{a,t} + \varepsilon_{a,t}. \tag{2.5}$$

The dependent variable $R_{a,t}$ is the monthly return to a long-short portfolio that is based on the extreme deciles of anomaly a. Post-Sample (S) is equal to one if the month is after the sample period used in the original study but still pre-publication and zero otherwise. Post-Publication (P) is equal to one if the month is after the official publication date and zero otherwise. Standard errors (in parentheses) are computed under the assumption of contemporaneous cross-sectional correlation between panel portfolio residuals.

Table 2.4 here.

Table 2.4 presents the results of the regression in Equation 2.5. As we can see, the coefficients of the Post-Sample dummy and the Post-Publication dummy are all negative like with McLean and Pontiff [2016]. But unlike with McLean and Pontiff [2016], the coefficient of Post-Sample dummy is not significant. The conclusion is the same as with McLean and Pontiff [2016]: when an anomaly is revealed to the world, it destroys the return of strategies based on this anomaly.

2.4.3 Forecasting the Return of Anomaly-based Strategies

a. Forecasting the Post-publication Returns

Table 2.5 presents the results of Equation 2.3 with the tone index being measured by the percentage of positive words and by the difference between the percentage of positive words and the percentage of negative words. I also add the year fixed effect and the anomaly fixed effect. Results with the tone being measured by negative words are not presented because it was inconclusive. The year fixed effect is important here because the impact factor of most of the journals considered in this paper tends to increase year after year.

Table 2.5 here.

In Table 2.5, the tone index is measured by the percentage of positive words in the abstract for columns (1) to (3). For columns (4) to (6), the tone index is measured by the difference between the percentage of positive words and the percentage of negative words in the abstract. As we can see, for all the models, the coefficient of the tone is positive and significant. The coefficient of the journal impact factor is positive and significant only if the tone is measured by the percentage of positive words. For all the models, the coefficient of the interaction is negative and significant. Taking into account the fixed effects does not change significantly the coefficients.

Let us consider the first three columns where the tone is measured by the percentage of positive words. To see the effect of the tone on the post publication return the expression $\beta_1 + \beta_3.JIF$ is the one to be considered. A negative $\beta_1 + \beta_3.JIF$ means an increase of the publication tone has a negative effect on the returns (the returns of strategies based on anomalies are destroyed). Therefore, a positive tone publication that discusses a given anomaly and appears in a high impact factor journal will tend to destroy the return of strategies based on this anomaly. As a matter of fact, since β_1 is positive and β_3 negative, for the expression $\beta_1 + \beta_3.JIF$ to be negative, JIF needs to be high enough. By dividing β_1 by β_3, the threshold of the JIF is obtained. By using the values of coefficients in Table 2.5, JIF needs to be between 2.5 and 3. A positive tone publication that appears in a journal with an impact factor higher than 3, is more likely to destroy the returns of strategies based on anomalies discussed in this publication.

The same conclusion is drawn if we consider the last three columns where the tone is measured by the difference between the percentage of positive words and the percentage of negative words. Values of JIF that make a positive tone publication destroy the returns on anomalies are between 2.4 and 3.2. To sum up, when an anomaly is discussed in a positive tone publication that appear in a journal with an impact factor higher than 3, this anomaly will attract investors that are going to arbitrage away the mispricing and therefore destroy the anomaly.

b. Forecasting the Post-publication Average Return

Table 2.6 and Table 2.7 present the results of Equation 2.4 with the tone index being measured by the number of positive words in Table 2.6 and by the difference between the number of positive words and the number of negative words in Table 2.7. Like said above, results with the tone being measured by negative words are not presented because it was inconclusive. As expected, the fitness of the model gets better as q increase. This can be observed in Figure 2.3. As we can see the scatter seems to adapt better to the regression line as q increases. q is the number of months separating a month in which a given anomaly is discussed in a publication from the next month this anomaly is discussed in a publication.

Figure 2.3 and Table 2.6 here.

In Table 2.6, as we can see, coefficients become significant for $q = 4$. The coefficient $beta_1$ and $beta_2$ which are the coefficients of the tone index and the journal impact factor respectively are positive while the coefficient of the interaction $beta_3$ is negative. A positive tone index causes destruction of the anomaly if $\beta_1+\beta_3.JIF_{a,m_i}$ is negative. Given that $\beta_1 > 0$ and $\beta_3 < 0$, this will happen if the journal impact factor is high enough to make the value $\beta_1 + \beta_3.JIF_{a,m_i}$ negative. This value is negative for journal impact factors higher than 3.5. In conclusion, an anomaly that got discussed in positive tone publications that appear in high impact factor journals has high chances of being destroyed.

Table 2.7 here.

The same conclusion goes for Table 2.7. In this table where I use the difference between the number of positive words and the number of negative words as the tone index, coefficients start being significant from $q = 5$. The interaction coefficient start being significant from $q = 15$. From $q = 15$, $\beta_1 + \beta_3.JIF_{a,m_i}$ is negative for values of journal impact factor higher than 3.

Table 2.8 and Table 2.9 present the results of Equation 2.4 with the tone index being measured by the percentage of positive words in Table 2.8 and by the difference between the percentage of positive words and the percentage of negative words in Table 2.9. In Table 2.8 coefficients become significant from $q = 4$. Coefficients of the interaction are not significant. The coefficient of the journal impact factor is only significant for $q = 4$ and $q = 8$. In Table 2.9, coefficients become significant from $q = 6$. The coefficient of the tone index is positive and significant from $q = 8$ and the coefficient of the interaction is negative and significant from $q = 15$. The coefficient of the journal impact factor is not significant. From $q = 15$, $\beta_1 + \beta_3.JIF_{a,m_i}$ is negative for values of journal impact factor higher than 4.

Table 2.8 and Table 2.9 here.

For all the tables analyzed in this part, the conclusion is quite the same: the longer the period separating the month of a publication discussing a given anomaly from the next month this anomaly appears in a publication, the more significant the impact of such publication.

2.5 Robustness Checks

In this section, some robustness checks are done. To make sure the results obtained are robust, this paper checks if the results still hold if only publications from the *Journal of*

Banking and Finance are considered since numerous publications about anomalies appear in this journal. It also checks what happens to the results if the transaction costs are considered.

2.5.1 Effect of Publications from Journal of Banking & Finance

To robustify its contribution in the sense of showing that it is tone plus the impact factor of the journal that provide the results, this paper runs the regression in Equation 2.3 by considering only the publications about anomalies that appear in the *Journal of Banking and Finance*. In fact, out of the 959 articles (that discuss at least an anomaly) considered in this paper, 132 appear in the Journal of banking and Finance, which represents approximately 14%. Meanwhile, the impact factor of the *Journal of Banking and Finance* through the years does not exceed 2.205. Given that results show that a publication with a positive tone has an impact on the behaviour of investors when it appears in a publication with an impact factor between 2.4 and 3.2, I have to make sure this result really holds by running the regression only with articles that appear in the *Journal of Banking and Finance*.

Table 2.10 here.

Table 2.10 presents the results of Equation 2.3 with articles that appear only in the *Journal of Banking and Finance*. As we can see, the coefficients of the tone index, the journal impact factor and the interaction between the tone index and the journal impact factor are no more significant, unlike when all the publications are considered. This confirms the fact that a publication with a high tone index needs to appear in a famous journal in order to draw investors' attention.

2.5.2 Effect of Transaction Costs

In this subsection, I do another robustness check by taking in account the transaction costs. Taking advantage of anomalies implies building portfolios that need to be rebalanced at a certain frequency. This means incurring transaction costs at each rebalancing point. Transaction costs could be a limit to arbitrage (Shleifer and Vishny [1997]) and deter investors from taking advantage of an anomaly even if positive things are published about this anomaly in the journals. I take the regression in Equation 2.3, and add to the independent variables, the average transaction cost of the stocks in each anomaly-based long-short portfolio. The equation to be estimated is the following:

$$R_{a,t} = \alpha + \beta_1.TI_{p_{a,t}} + \beta_2.JIF_{a,t} + \beta_3.TI_{p_{a,t}}.JIF_{a,t} + \beta_4.TCOST_{a,t} + \varepsilon_{a,t}, \tag{2.6}$$

where $R_{a,t}$ is the return of long-short strategy based on anomaly a at month t. $TI_{p_{a,t}}$ is the average tone index of the latest publications in which anomaly a has been discussed and $JIF_{a,t}$ is the average impact factor of the journals in which the latest publications that discuss anomaly a appear. By *latest publications* here, I mean those that appear during the latest of the months t' previous to t, $t' < t$. $TCOST_{a,t}$ is the average transaction cost of the stocks in the long-short portfolio based on anomaly a at month t. This portfolio contains the stocks in the first and tenth decile stocks are ranked according to anomaly a. $\varepsilon_{a,t}$ is the error term.

The transaction cost is computed according to Hasbrouck [2009] with an extension to incorporate funding liquidity as a determinant of transaction cost. In fact, transaction costs are high when funding conditions are tight (Brunnermeier and Pedersen [2009]. The TED spread is used as a measure of funding liquidity. The transaction cost calculation is presented in Appendix 2.6.

Table 2.11 here.

The results of the regression are presented in Table 2.11. As we can see, the coefficient of the transaction cost is positive and significant. This means that when the transaction costs of the stocks composing the anomaly-based long-short portfolio increases, the anomaly return increases. This is explained by the fact that a high transaction cost prevents investors from taking advantage of anomalies and arbitrage away the mispricing due to anomalies. The conclusion about the tone index and the journal impact factor still holds: a publication with a high tone index needs to appear in a famous journal (journal impact factor higher than 3) in order to draw the attention of the investors. Results stay the same when an anomaly fixed-effect is introduced. But when the year fixed effect is introduced, the conclusion about the tone index and the journal impact factor still holds but the coefficient of the transaction is no more significant. This is due to the fact that transaction costs are decreasing as time goes by, and there is less variability for the transaction costs of anomaly-based portfolio considered within a giving year.

2.6 Conclusion

Anomalies suggest that an asset could be mispriced and may provide some opportunities to arbitrageurs. To take advantage of this mispricing, arbitrageurs can use anomaly-based long-short portfolios that need to be rebalanced frequently. For every anomaly, once it got discovered, academic research usually write a lot about it to explain why it is there, when or in which conditions it ceases to be there. This paper assesses the publication impact on returns of strategies based on anomalies not by focusing on the first publication on the anomaly but on all the publications on the anomaly.

This paper extracts the tone from the abstract of all the academic publications in which anomalies appear and constructs a tone index. The tone index and the journal impact factor in which the anomalies are published are used to forecast the returns of long-short anomaly-based portfolios. The dependant variables are the post-publication returns and the average return of the anomaly-based portfolio over the period going from the month the anomaly is discussed in a publication to the next month it is discussed again in a publication. Results show that when an anomaly is discussed in a positive tone publication that appears in a high impact factor journal, the return of the long-short portfolio based on this anomaly decreases. Anomalies are therefore only destroyed when they appear in high-impact-factor journals. This paper provides evidence that publications in top finance journals are more trustworthy for investors.

Tables and Figures of Chapter 2

Table 2.1: **Summary Statistics by Journal**

Journals	**Number of publications**	**Percentage of publications**	**Number of pub with anomalies**
Accounting Review	1465	2,18%	55
Econometrica	1562	2,33%	3
Financial Analysts Journal	605	0,90%	50
Journal of Accounting & Economics	914	1,36%	31
Journal of Accounting Research	629	0,94%	27
Journal of Banking & Finance	4621	6,89%	132
Journal of Business Finance & Accounting	767	1,14%	37
Journal of Econometrics	3303	4,93%	11
Journal of Finance	2180	3,25%	118
Journal of Financial and Quantitative Analysis	1315	1,96%	88
Journal of Financial Econometrics	276	0,41%	4
Journal of Financial Economics	2421	3,61%	162
Journal of Financial Intermediation	536	0,80%	8
Journal of Financial Markets	442	0,66%	28
Journal of Financial Research	242	0,36%	11
Journal of International Financial Management	128	0,19%	2
Journal of Investment Management	81	0,12%	-
Journal of Money Credit and Banking	1550	2,31%	9
Journal of Political Economy	1215	1,81%	6
Management Science	5039	7,51%	42
Quarterly Journal of Economics	1171	1,75%	6
Review of Accounting Studies	584	0,87%	37
Review of Financial Studies	1932	2,88%	92
TOTAL	**32,978**	**100.00%**	959

Table 2.2: **Anomalies and Average Tone index**

Anomalies	**Nbr of pub**	**Nbrpos**	**Nbrneg**	**%pos**	**%neg**	**TI (percentage of words)**
Indust momentum	18	7.36	6.50	5.80%	5.45%	0.35%
Momentum	88	6.61	4.94	5.43%	3.58%	1.85%
Value	111	7.16	10.70	5.67%	6.44%	-0.77%
Size	66	7.55	5.23	5.90%	4.09%	1.81%
Accrual	34	5.16	10.00	4.38%	6.67%	-2.28%
SUE	109	7.86	5.26	6.50%	4.29%	2.20%
Asset growth	41	3.50	9.79	3.96%	7.79%	-3.84%
Share issuance	60	7.64	5.40	7.23%	4.49%	2.74%
Low-volatility	4	3.41	11.11	2.51%	9.09%	-6.58%
Long-term reversal	15	15.06	2.67	9.95%	4.92%	5.03%
Seasonality	15	4.96	11.33	3.27%	7.40%	-4.13%
Investment to K	2	12.83	5.29	8.28%	4.08%	4.20%
Earnings/Price	39	6.03	7.28	4.87%	5.56%	-0.69%
Cashflow/Price	6	10.05	13.50	8.32%	7.26%	1.06%
Market Beta	47	6.71	15.00	5.10%	7.46%	-2.36%
Idio-vol	28	11.75	8.98	9.00%	6.15%	2.85%
Short-term reversal	15	3.67	11.04	4.58%	5.05%	-0.47%
Dividend Yield	144	10.50	7.61	11.74%	4.84%	6.89%
Gross Profitability	56	5.00	13.97	5.77%	7.93%	-2.16%
Return on Asset	6	10.00	2.00	6.44%	0.87%	5.58%
Sales Growth	30	5.79	22.00	4.54%	9.52%	-4.99%
Gross Margins	1	8.14	8.80	6.25%	5.47%	0.78%
Asset Turnover	6	9.00	13.17	6.52%	7.52%	-1.00%

Table 2.3: **Anomalies and JIF**

Anomalies	**First Publications**	**Number of publications**	**Average JIF of journals of publication**
Industry momentum	Moskowitz and Grinblatt [1999a]	18	2.08
Momentum	Jegadeesh and Titman [1993b]	88	2.14
Value	Fama and French [1993]	111	2.74
Size	Fama and French [1993]	66	2.33
Accrual	Sloan [1996]	34	2.04
Standard unexpected earnings	Foster et al. [1984]	109	2.30
Asset growth	Cooper et al. [2008]	41	2.66
Share issuance	Pontiff and Woodgate [2008]	60	3.07
Low-volatility	Ang et al. [2006a]	4	1.07
Long-term reversal	De Bondt and Thaler [1985]	15	2.67
Seasonality	Heston and Sadka [2008]	15	2.26
Investment to Capital	Xing [2007]	2	3.09
Earnings/Price	Basu [1977]	39	1.98
Cashflow/Price	Chan et al. [1991]	6	2.46
Market Beta	Fama and MacBeth [1973]	47	2.77
Idiosyncratic Volatility	Ang et al. [2006b]	28	2.50
Short-term reversal	Jegadeesh [1990]	15	1.67
Dividend Yield	Naranjo et al. [1998]	144	2.88
Gross Profitability	Novy-Marx [2013]	56	2.64
Return on Asset	Chen et al. [2011]	6	2.25
Sales Growth	Lakonishok et al. [1994]	30	2.87
Gross Margins	Novy-Marx [2013]	1	0.81
Asset Turnover	Soliman [2008]	6	2.64

Table 2.4: **Regression of portfolio Returns on post-sample dummy and on post-publication dummy**

The regressions test for changes in returns relative to the anomaly's sample-end and publication dates. The regression equation is the following:

$R_{a,t} = \alpha_i + \beta_1$.Post Sample Dummy$_{a,t} + \beta_2$.Post Publication Dummy$_{a,t} + \varepsilon_{a,t}$.

The dependent variable $R_{a,t}$ is the monthly return to a long-short portfolio that is based on the extreme deciles of anomaly a. Post-Sample (S) is equal to one if the month is after the sample period used in the original study but still pre-publication and zero otherwise. Post-Publication (P) is equal to one if the month is after the official publication date and zero otherwise. Standard errors (in parentheses) are computed under the assumption of contemporaneous cross-sectional correlation between panel portfolio residuals. *, **, and *** denote statistical signicance at the 10%, 5%, and 1% levels, respectively.

Anomaly FE : YES

Sample size: 9729

	Post-Sample (S)	**Post-Publication (P)**
Coefficients	-0.156	-0.317 ***
Errors	(0.092)	(0.058)

Table 2.5: **Regression of post publication returns on publication tone and journal impact factor**

The regression tests for effects of publications on anomalies in peer-reviewed journal on the return of the anomaly following the months of the publication. The dependent variable is the average of the monthly return to a long-short portfolio that is based on the extreme deciles of anomaly. It is a panel regression. The equation of the regression is the following:

$$R_{a,t} = \alpha + \beta_1.TI_{p_{a,t}} + \beta_2.JIF_{a,t} + \beta_3.TI_{p_{a,t}}.JIF_{a,t} + \varepsilon_{a,t}.$$

$R_{a,t}$ is the return of long-short strategy based on anomaly a at month t. $TI_{p_{a,t}}$ is the average tone index of the latest publications in which anomaly a has been discussed and $JIF_{a,t}$ is the average impact factor of the journals in which the latest publications that discuss anomaly a appear. By *latest publications* here, I mean those that appear during the latest of the months t' previous to t, $t' < t$. $\varepsilon_{a,t}$ is the error term. The estimations of coefficients α, β_1, β_2 and β_3 are presented for each column. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are in parentheses. In columns (1) to (3), the tone index is measured by the percentage of positive words in the abstract. In columns (4) to (6), the tone index is measured by the difference between the percentage of positive words and the percentage of negative words in the abstract.

	Tone = %pos			**Tone = %pos -%neg**		
Covariates	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Tone (β_1)	15.517 ***	12.368 **	16.823 ***	10.279 ***	9.105 ***	8.938 **
	(5.699)	(5.338)	(5.915)	(3.593)	(3.325)	(4.073)
JIF (β_2)	0.312 **	0.395 ***	0.314 **	-0.019	0.037	-0.005
	(0.123)	(0.117)	(0.128)	(0.083)	(0.087)	(0.085)
Tone.JIF (β_3)	-5.400 **	-5.017 **	-5.778 ***	-3.437 **	-3.844 ***	-2.792 *
	(2.168)	(2.083)	(2.223)	(1.412)	(1.342)	(1.532)
Intercept (α)	-0.629 **	-0.635	-1.296 ***	0.321	0.154	-0.321
	(0.319)	(0.749)	(0.490)	(0.227)	(0.735)	(0.427)
Year fixed effect	No	Yes	No	No	Yes	No
Anomaly fixed effect	No	No	Yes	No	No	Yes
Sample size	6,072	6,072	6,072	6,072	6,072	6,072

Table 2.6: **Regression of average post publication return on publication tone (number of positive words minus number of negative words in publication abstract and journal impact factor**

The regressions test for effects of publications on anomalies in peer-reviewed journal on the average return of the anomaly over the period following the publication. The dependent variable is the average of the monthly return to a long-short portfolio that is based on the extreme deciles of anomaly. The equation of the regression is the following:

$R_{a,[m_i,m_{i+1}]} = \alpha + \beta_1.Nbrpos_{a,m_i} + \beta_2.JIF_{a,m_i} + \beta_3.Nbrpos_{a,m_i}.JIF_{a,m_i} + \varepsilon_{a,m_i}.$

$[m_i, m_{i+1}]$ represent the interval between two successive months of appearance of anomaly a in a publication. It means none of the publications that appears between m_i, m_{i+1} discuses anomaly a. $R_{a,[m_i,m_{i+1}]}$ is the average return of a long-short strategy based on anomaly a over the period $[m_i, m_{i+1}]$. $Nbrpos_{a,m_i}$ is the average number of positive words in publications that discussed anomaly a during month m_i. JIF_{a,m_i} is the average impact factor of journals in which publications that discussed anomaly a appear during month m_i. The first column represent q the minimum number of months separating month m_i from the next month m_{i+1}. The estimations of coefficients α, β_1, β_2 and β_3 are presented in the following columns. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The adjusted R^2 is presented in the sixth column and the sample size is presented in the last column.

q	**Intercept** α	**Nbrpos** β_1	**JIF** β_2	**Nbrpos.JIF** β_3	**Adj.Rsq**	**Sample size**
1	0.502	0.057	-0.068	-0.025	-0.02%	750
2	0.291	0.052	0.107	-0.040	-0.03%	544
3	-0.154	0.029	0.277	-0.036	-0.13%	379
4	-0.674	0.159 **	0.352 *	-0.052	1.21%	292
5	-0.513	0.145 **	0.218	-0.047	1.02%	240
6	-0.394	0.166 **	0.150	-0.044	3.81%	199
7	-0.412	0.139 **	0.153	-0.034	3.07%	170
8	-1.162 **	0.199 ***	0.450 **	-0.057 **	7.91%	146
9	-0.992 *	0.188 **	0.321	-0.049 *	8.16%	129
10	-1.209 **	0.206 ***	0.392	-0.055 **	10.06%	121
11	-1.190 **	0.193 **	0.320	-0.047 *	9.56%	113
12	-1.377 **	0.252 **	0.392	-0.072 *	13.91%	95
13	-1.049 *	0.199 **	0.345	-0.061 *	8.90%	80
14	-0.940	0.200 **	0.324	-0.062 *	10.03%	76
15	-0.965	0.206 **	0.340	-0.064 *	12.23%	69
16	-1.005	0.209 **	0.359	-0.065 *	11.55%	62
17	-0.997	0.206 **	0.368	-0.067 *	9.78%	61
18	-0.986	0.206 **	0.370	-0.067 *	9.38%	59

Table 2.7: **Regression of average post publication return on publication tone and journal impact factor**

The regressions test for effects of publications on anomalies in peer-reviewed journal on the average return of the anomaly over the period following the publication. The dependent variable is the average of the monthly return to a long-short portfolio that is based on the extreme deciles of anomaly. The equation of the regression is the following:

$$R_{a,[m_i,m_{i+1}]} = \alpha + \beta_1.(Nbrpos_{a,m_i} - Nbrneg_{a,m_i}) + \beta_2.JIF_{a,m_i} + \beta_3.(Nbrpos_{a,m_i} - Nbrneg_{a,m_i}).JIF_{a,m_i} + \varepsilon_{a,m_i}.$$

$[m_i, m_{i+1}]$ represent the interval between two successive months of appearance of anomaly a in a publication. It means none of the publications that appears between m_i, m_{i+1} discuses anomaly a. $R_{a,[m_i,m_{i+1}]}$ is the average return of a long-short strategy based on anomaly a over the period $[m_i, m_{i+1}]$. $Nbrpos_{a,m_i}$ is the average number of positive words and $Nbrneg_{a,m_i}$ is the average number of negative words in publications that discussed anomaly a during month m_i. JIF_{a,m_i} is the average impact factor of journals in which publications that discussed anomaly a appear during month m_i. The first column represent q the minimum number of months separating month m_i from the next month m_{i+1}. The estimations of coefficients α, β_1, β_2 and β_3 are presented in the following columns. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The adjusted R^2 is presented in the sixth column and the sample size is presented in the last column.

q	**Intercept**	**Nbrpos-Nbrneg**	**JIF**	**(Nbrpos-Nbrneg).JIF**	**Adj.Rsq**	**Sample size**
	α	β_1	β_2	β_3		
1	0.847 **	0.028	-0.220	-0.010	-0.16%	750
2	0.630	0.039	-0.176	-0.024	-0.27%	544
3	0.169	0.046	-0.008	-0.021	-0.73%	379
4	0.511	0.066	-0.056	-0.027	-0.07%	292
5	0.601	0.073 *	-0.166	-0.032	0.15%	240
6	0.910 **	0.084 **	-0.202	-0.026	2.00%	199
7	0.682 *	0.073 *	-0.138	-0.024	1.45%	170
8	0.377	0.089 **	-0.021	-0.029	5.34%	146
9	0.426	0.079 *	-0.055	-0.020	4.43%	129
10	0.345	0.080 *	-0.029	-0.020	4.67%	121
11	0.206	0.070	-0.015	-0.016	2.81%	113
12	0.483	0.079	-0.138	-0.019	4.32%	95
13	0.384	0.068	-0.113	-0.021	0.27%	80
14	0.524	0.072	-0.154	-0.025	1.04%	76
15	0.594	0.141 ***	-0.172	-0.048 **	15.06%	69
16	0.580	0.141 ***	-0.164	-0.048 **	13.66%	62
17	0.578	0.142 ***	-0.181	-0.050 **	13.76%	61
18	0.588	0.141 ***	-0.181	-0.050 **	13.47%	59

Table 2.8: **Regression of average post publication return on publication tone (%age of positive words) and journal impact factor**

The regressions test for effects of publications on anomalies in peer-reviewed journal on the average return of the anomaly over the period following the publication. The dependent variable is the average of the monthly return to a long-short portfolio that is based on the extreme deciles of anomaly. The equation of the regression is the following:

$$R_{a,[m_i,m_{i+1}]} = \alpha + \beta_1.(\%pos_{a,m_i}) + \beta_2.JIF_{a,m_i} + \beta_3.(\%pos_{a,m_i}).JIF_{a,m_i} + \varepsilon_{a,m_i}.$$

$[m_i, m_{i+1}]$ represent the interval between two successive months of appearance of anomaly a in a publication. It means none of the publications that appears between m_i, m_{i+1} discuses anomaly a. $R_{a,[m_i,m_{i+1}]}$ is the average return of a long-short strategy based on anomaly a over the period $[m_i, m_{i+1}]$. $\%pos_{a,m_i}$ is the average percentage of positive words in the abstracts of publications that discussed anomaly a during month m_i. JIF_{a,m_i} is the average impact factor of journals in which publications that discussed anomaly a appear during month m_i. The first column represent q the minimum number of months separating month m_i from the next month m_{i+1}. The estimations of coefficients α, β_1, β_2 and β_3 are presented in the following columns. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The adjusted R^2 is presented in the sixth column and the sample size is presented in the last column.

q	**Intercept** α	**%pos** β_1	**JIF** β_2	**(%pos).JIF** β_3	**Adj.Rsq**	**Sample size**
1	0.242	6.264	-0.072	-1.513	-0.18%	750
2	0.080	6.942	0.095	-3.290	-0.38%	544
3	-0.309	7.512	0.272	-4.604	-0.11%	379
4	-0.619	13.314	0.387 *	-6.833	0.49%	292
5	-0.283	5.943	0.222	-4.282	-0.45%	240
6	-0.811	19.829	0.365	-7.461	0.34%	199
7	-0.473	10.175	0.286	-4.788	-1.29%	170
8	-1.350 **	24.530 ***	0.519 **	-6.230	9.05%	146
9	-1.041 **	22.601 **	0.340	-4.724	8.51%	129
10	-1.136 **	26.245 ***	0.380	-5.932	9.76%	121
11	-1.131 **	26.006 **	0.265	-4.283	12.73%	113
12	-1.465 **	35.864 **	0.424	-8.732	14.41%	95
13	-1.075	28.053 **	0.391	-7.296	7.23%	80
14	-0.857	27.509 **	0.275	-6.724	7.23%	76
15	-0.997	31.685 **	0.329	-8.040	10.49%	69
16	-0.972	30.821 **	0.325	-7.674	9.46%	62
17	-0.913	31.949 **	0.234	-7.414	9.24%	61
18	-0.902	31.983 **	0.238	-7.490	8.92%	59

Table 2.9: **Regression of average post publication return on publication tone (%age of positive words minus %age of negative words) and journal impact factor**

The regressions test for effects of publications on anomalies in peer-reviewed journal on the average return of the anomaly over the period following the publication. The dependent variable is the average of the monthly return to a long-short portfolio that is based on the extreme deciles of anomaly. The equation of the regression is the following:

$$R_{a,[m_i,m_{i+1}]} = \alpha + \beta_1.(\%pos_{a,m_i} - \%neg_{a,m_i}) + \beta_2.JIF_{a,m_i} + \beta_3.(\%pos_{a,m_i} - \%neg_{a,m_i}).JIF_{a,m_i} + \varepsilon_{a,m_i}.$$

$[m_i, m_{i+1}]$ represent the interval between two successive months of appearance of anomaly a in a publication. It means none of the publications that appears between m_i, m_{i+1} discuses anomaly a. $R_{a,[m_i,m_{i+1}]}$ is the average return of a long-short strategy based on anomaly a over the period $[m_i, m_{i+1}]$. $\%pos_{a,m_i}$ is the average percentage of positive words and $\%neg_{a,m_i}$ is the average % of negative words in the abstracts of publications that discussed anomaly a during month m_i. JIF_{a,m_i} is the average impact factor of journals in which publications that discussed anomaly a appear during month m_i. The first column represent q the minimum number of months separating month m_i from the next month m_{i+1}. The estimations of coefficients α, β_1, β_2 and β_3 are presented in the following columns. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The adjusted R^2 is presented in the sixth column and the sample size is presented in the last column.

q	**Intercept**	**%pos-%neg**	**JIF**	**(%pos-%neg).JIF**	**Adj.Rsq**	**Sample size**
	α	β_1	β_2	β_3		
1	0.418	-1.509	-0.103	0.619	-0.32%	750
2	0.409	2.543	-0.094	-1.939	-0.30%	544
3	0.058	3.241	0.004	-2.574	0.09%	379
4	-0.023	1.645	0.012	-2.438	0.69%	292
5	0.126	3.477	-0.091	-3.476	0.66%	240
6	0.390	10.972	-0.136	-5.598 **	0.70%	199
7	0.313	10.504	-0.123	-6.067 *	0.73%	170
8	0.071	13.392 **	0.139	-3.746	6.93%	146
9	0.244	11.995 **	0.075	-2.443	6.48%	129
10	0.392	14.906 **	0.030	-3.382	8.86%	121
11	0.232	12.701 **	0.071	-2.053	9.37%	113
12	0.516	16.166 **	-0.043	-3.366	11.12%	95
13	0.455	14.813 **	0.003	-3.231	8.84%	80
14	0.692	15.679 **	-0.100	-3.354	9.85%	76
15	1.014 **	29.190 ***	-0.192	-7.285 **	28.54%	69
16	0.986 **	30.286 ***	-0.181	-7.186 **	28.72%	62
17	1.152 **	32.321 ***	-0.300	-8.466 **	30.25%	61
18	1.163 **	32.269 ***	-0.300	-8.470 **	29.98%	59

Table 2.10: **Regression of post publication returns on publication tone and journal impact factor considering only Journal of Banking and Finance**

The regression tests for effects of publications on anomalies in Journal of Banking and Finance on the return of the anomaly following the months of the publication. The dependent variable is the average of the monthly return to a long-short portfolio that is based on the extreme deciles of anomaly. It is a panel regression. The equation of the regression is the following:

$R_{a,t} = \alpha + \beta_1.TI_{p_{a,t}} + \beta_2.JIF_{a,t} + \beta_3.TI_{p_{a,t}}.JIF_{a,t} + \varepsilon_{a,t}.$

$R_{a,t}$ is the return of long-short strategy based on anomaly a at month t. $TI_{p_{a,t}}$ is the average tone index of the latest publications in which anomaly a has been discussed and $JIF_{a,t}$ is the average impact factor of the journals in which the latest publications that discuss anomaly a appear. By *latest publications* here, I mean those that appear during the latest of the months t' previous to t, $t' < t$. $\varepsilon_{a,t}$ is the error term. The estimations of coefficients α, β_1, β_2 and β_3 are presented for each column. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are in parentheses. In columns (1) to (3), the tone index is measured by the percentage of positive words in the abstract. In columns (4) to (6), the tone index is measured by the difference between the percentage of positive words and the percentage of negative words in the abstract.

	Tone = %pos			**Tone = %pos -%neg**		
Covariates	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Tone (β_1)	28.058	40.029 *	32.451	17.957	28.710 *	16.812
	(20.098)	(22.583)	(22.641)	(15.280)	(16.798)	(16.897)
JIF (β_2)	0.064	0.885	-0.097	-0.343	-0.051	-0.642
	(0.779)	(1.540)	(0.913)	(0.439)	(1.347)	(0.530)
Tone.JIF (β_3)	-8.618	-13.813	-11.947	-8.058	-13.351	-4.345
	(13.227)	(14.748)	(14.631)	(9.573)	(10.577)	(10.885)
Intercept (α)	-0.978	-12.303 ***	-0.725	0.446	-10.231 **	0.920
	(1.173)	(4.487)	(1.409)	(0.659)	(4.374)	(0.777)
Year fixed effect	No	Yes	No	No	Yes	No
Anomaly fixed effect	No	No	Yes	No	No	Yes
Sample size	584	584	584	584	584	584
Number of anomalies	16	16	16	16	16	16

Table 2.11: **Regression of post publication returns on transaction cost, publication tone and journal impact factor considering only Journal of Banking and Finance**

The regression tests for effects of publications on anomalies in peer-reviewed journal on the return of the anomaly following the months of the publication. The dependent variable is the average of the monthly return to a long-short portfolio that is based on the extreme deciles of anomaly. It is a panel regression. The equation of the regression is the following:

$$R_{a,t} = \alpha + \beta_1.TI_{p_{a,t}} + \beta_2.JIF_{a,t} + \beta_3.TI_{p_{a,t}}.JIF_{a,t} + \beta_4.TCOST_{a,t} + \varepsilon_{a,t}.$$

$R_{a,t}$ is the return of long-short strategy based on anomaly a at month t. $TI_{p_{a,t}}$ is the average tone index of the latest publications in which anomaly a has been discussed and $JIF_{a,t}$ is the average impact factor of the journals in which the latest publications that discuss anomaly a appear. By *latest publications* here, I mean those that appear during the latest of the months t' previous to t, $t' < t$. $\varepsilon_{a,t}$ is the error term. The estimations of coefficients α, β_1, β_2 and β_3 are presented for each column. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are in parentheses. In columns (1) to (3), the tone index is measured by the percentage of positive words in the abstract. In columns (4) to (6), the tone index is measured by the difference between the percentage of positive words and the percentage of negative words in the abstract.

	Tone = %pos			**Tone = %pos -%neg**		
Covariates	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Tone (β_1)	14.782 ***	12.390 **	16.468 ***	9.378 ***	9.337 ***	8.311 **
	(5.632)	(5.339)	(5.913)	(3.569)	(3.338)	(4.079)
JIF (β_2)	0.352 ***	0.395 ***	0.352 ***	0.012	0.035	0.022
	(0.122)	(0.117)	(0.129)	(0.083)	(0.087)	(0.086)
Tone.JIF (β_3)	-5.455 **	-5.008 **	-5.910 ***	-3.374 **	-3.912 ***	-2.774 *
	(2.151)	(2.083)	(14.631)	(1.401)	(1.345)	(1.531)
Tcost (β_4)	0.222 ***	-0.102	0.194 **	0.208 ***	-0.143	0.185 **
	(0.077)	(0.179)	(0.079)	(0.078)	(0.178)	(0.079)
Intercept (α)	-1.328 ***	-0.115	-1.218 ***	-0.387	-0.548	-0.296
	(0.397)	(1.178)	(0.404)	(0.347)	(0.694)	(0.342)
Year fixed effect	No	Yes	No	No	Yes	No
Anomaly fixed effect	No	No	Yes	No	No	Yes
Sample size	6072	6072	6072	6072	6072	6072

Figure 2.1: **Evolution of Number of publications that discuss anomalies**

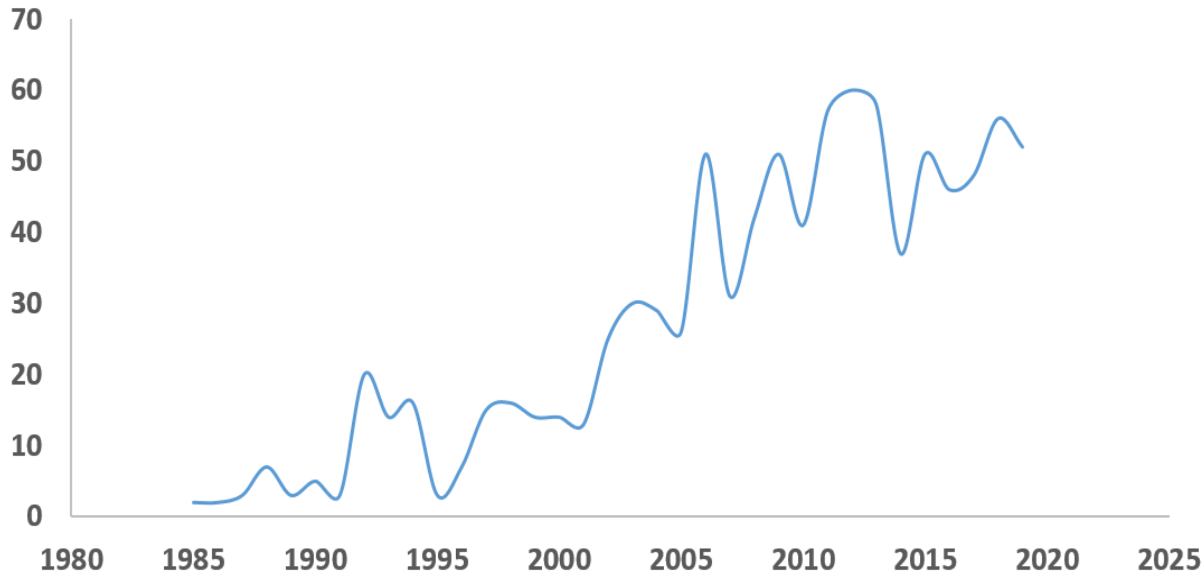


Figure 2.2: **Evolution of Percentage of publications that discuss anomalies**

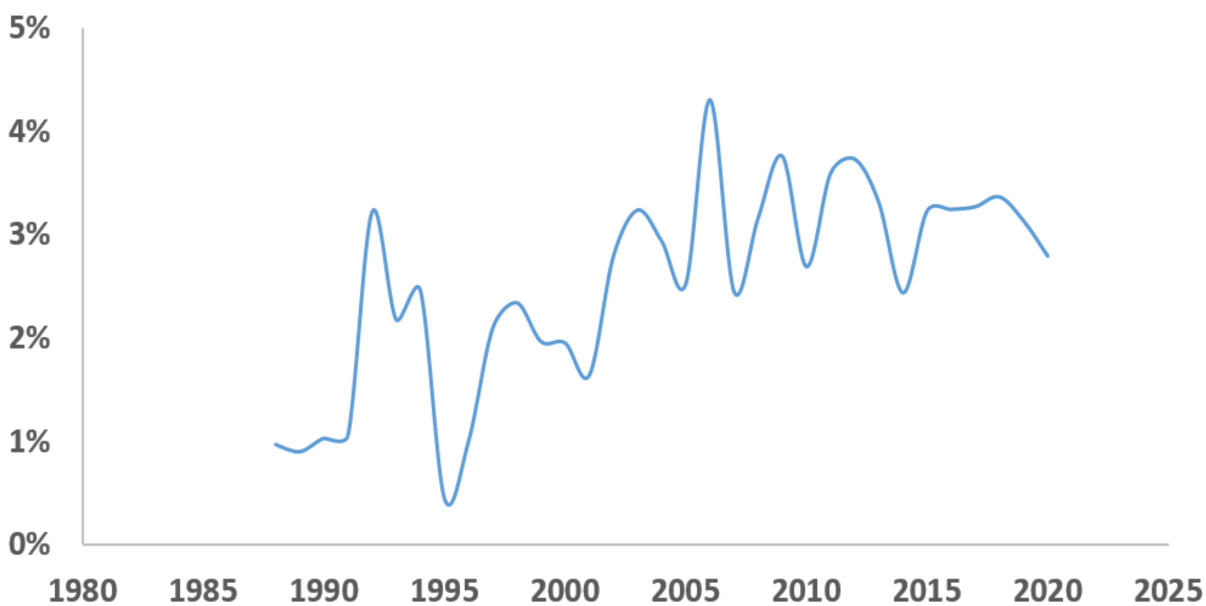


Figure 2.3: **Scatter plot between post publication average anomaly return and publication tone**

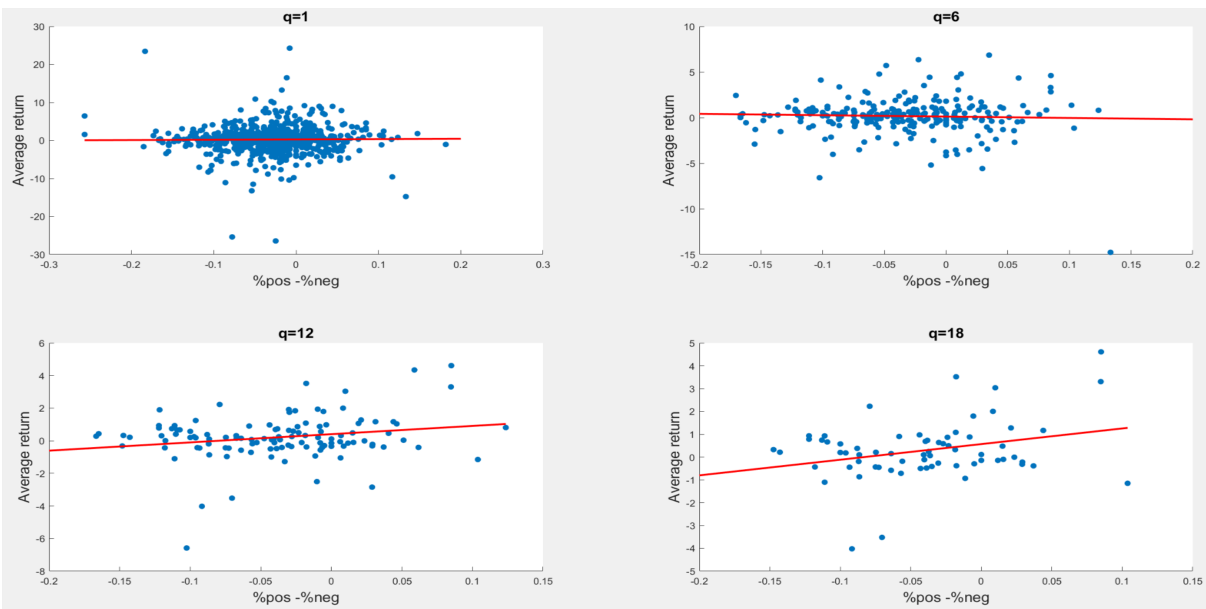
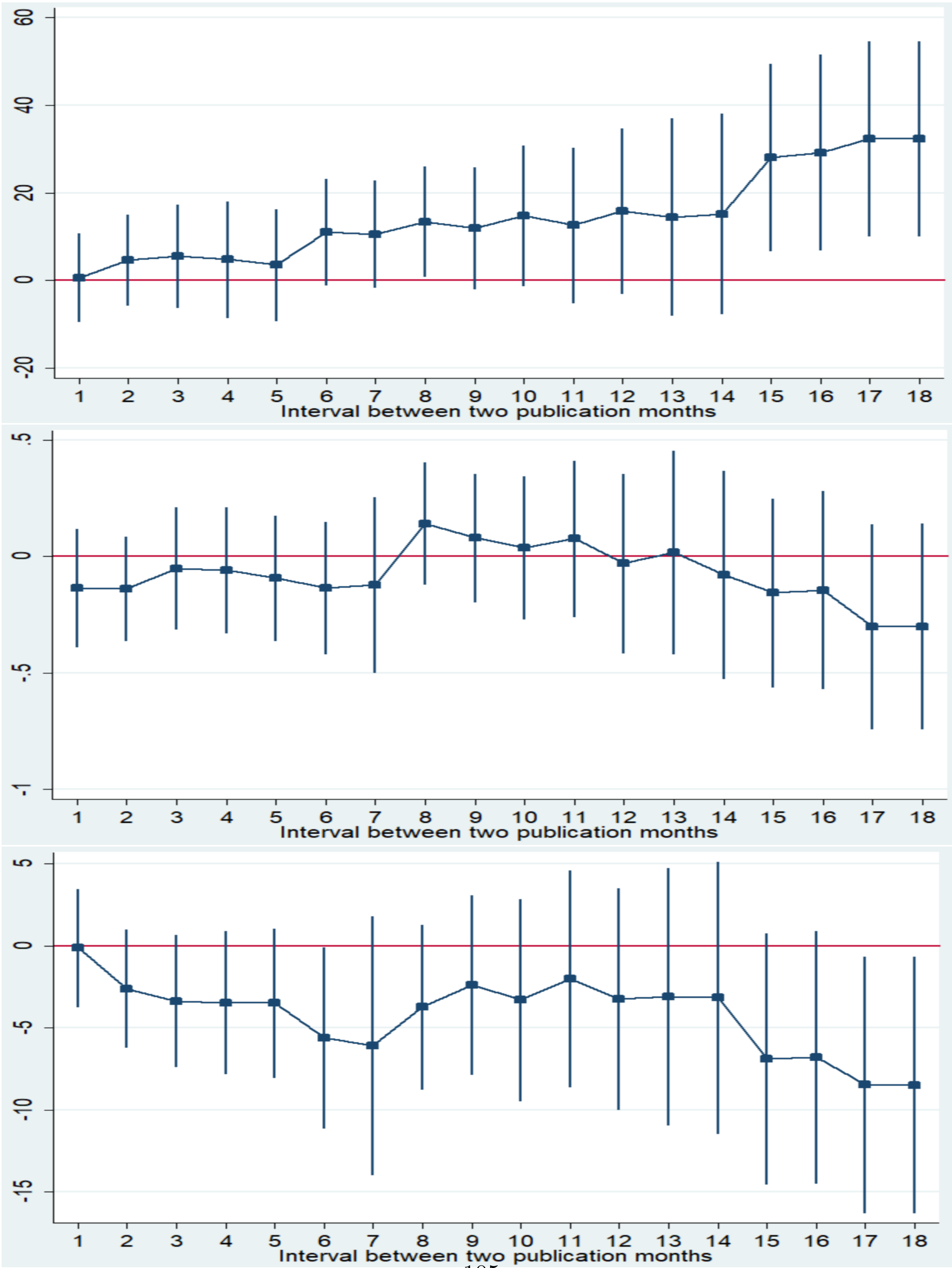


Figure 2.4: **Evolution of** $beta_1$**,** $beta_2$ **and** $beta_3$ **with respect to** q



Appendices for Chapter 2 (B)

B1 - Anomalies

The anomalies definitions and descriptions are based on the lists of characteristics compiled by Novy-Marx and Velikov [2016] and Kozak et al. [2019].

- **Industry Momentum (INDMOM)**:

  Follows Moskowitz and Grinblatt [1999b]. $INDMOM = rank(\sum_{l=1}^{6} r_{t-l}^{ind})$. In each month, the Fama and French 49 industries are ranked on their value-weighted past 6-months performance. Rebalanced monthly.

- **Momentum**:

  Follows Jegadeesh and Titman [1993a]. $MOM = \sum_{l=2}^{12} r_{t-l}$. Cumulated past performance in the previous 11 months by skipping the most recent month. Rebalanced monthly.

- **Size (SIZE)**:

  Follows Fama and French [1993]. $SIZE = ME_{Jun}$. We use the CRSP end of June price times shares outstanding. Updated annually.

- **Book-to Market (VALUE)**: The log of book value of equity scaled by market value of equity. Updated annually.

  Follows Fama and French [1993].

- **Accruals (ACC)**:

Follows Sloan [1996].

$$ACC = \frac{\Delta ACT - \Delta CHE - \Delta LCT + \Delta DLC + \Delta TXP - \Delta DP}{(AT + AT_{-12})/2}$$

, where ΔACT is the annual change in total current assets, ΔCHE is the annual change in total cash and short-term investments, ΔLCT is the annual change in current liabilities, ΔDLC is the annual change in debt in current liabilities, ΔTXP is the annual change in income taxes payable, ΔDP is the annual change in depreciation and amortization, and $(AT + AT_{-12})/2$ is average total assets over the last two years. Rebalanced annually.

- **Standardized Unexpected Earnings (SUE)**:

  Follows Foster et al. [1984]. $SUE = \frac{IBQ - IBQ_{-12}}{\sigma_{IBQ_{-24}:IBQ_{-3}}}$, where IBQ is income before extraordinary items (updated quarterly), and $\sigma_{IBQ_{-24}:IBQ_{-3}}$ is the standard deviation of IBQ in the past two years skipping the most recent quarter. Earnings surprises are measured by Standardized Unexpected Earnings (SUE), which is the change in the most recently announced quarterly earnings per share from its value announced four quarters ago divided by the standard deviation of this change in quarterly earnings over the prior eight quarters. Rebalanced monthly.

- **Asset Growth (AG)**:

  Follows Cooper et al. [2008]. $AG = AT/AT_{-12}$. Rebalanced annually.

- **Share Issuance (annual) (NISSA)**:

  Follows Pontiff and Woodgate [2008]. $NISSA = shrout_{Jun}/shrout_{Jun-12}$, where $shrout$ is the number of shares outstanding. Change in real number of shares outstanding from past June to June of the previous year. Excludes changes in shares due to stock dividends and splits, and companies with no changes in $shrout$.

- **Realized Volatility (REALVOL)**:

  Follows Ang et al. [2006b]. $REALVOL = \frac{252}{N}\sum_{t=1}^{N} r_t^2$. N is the number of available returns for the stock for the given year. Rebalanced annually.

- **Long-term Reversals (LTREV)**:

  Follows De Bondt and Thaler [1985]. $LTREV = \sum_{l=13}^{60} r_{t-l}$. Cumulative returns from $t - 60$ to $t - 13$. Updated monthly.

- **Seasonality (SEASON)**:

  Follows Heston and Sadka [2008]. $SEASON = \sum_{l=1}^{5} r_{t-l\times 12}$. Average monthly return in the same calendar month over the last 5 years. As an example, the average return from prior Octobers is used to predict returns this October. The firm needs at least one year of data to be included in the sample. Updated monthly.

- **Investment-to-Capital (IK)**:

  Follows Xing [2007]. $IG = CAPX/PPENT$. Investment to capital is the ratio of capital expenditure ($CAPX$) over property, plant, and equipment ($PPENT$).

- **Earnings-to-Price (E/P)**:

  Follows Basu [1977]. Net income scaled by market value of equity. Updated annually.

- **Cashflow-to-Price (C/P)**:

  Follows Chan et al. [1991]. Net income plus depreciation and amortization, all scaled by market value of equity. Updated annually.

- **Market beta (BETA)**:

  Fama and MacBeth [1973]. Beta with respect to the CRSP equal-weighted return index. Estimated over the past 60 months. Updated monthly.

- **Idiosyncratic volatility (IDIOVOL)**:

  Follows Ang et al. [2006b]. The standard deviation of the residual from a regression of daily stock returns on the daily innovations of the Fama and French three-factor model using 60 days (minimum 20) of lagged returns. Returns are market value-weighted.

- **Short-term Reversal (STREV)**:

  Follows Jegadeesh [1990]. $STREV = r_{t-1}$. Return in the previous month. Updated monthly.

- **Dividend Yield (D/P)**:

  Follows Naranjo et al. [1998]. The dividend yield use to form portfolios in June of year t is the total dividends paid from July of t-1 to June of t per dollar of equity in June of t.

- **Gross Profitability (GPROF)**:

  Follows Novy-Marx [2013]. $GPROF = GP/AT$, where GP is gross profits and AT is total assets. Rebalanced annually.

- **Return on Assets (annual) (ROAA)**:

  Follows Chen et al. [2011]. $ROAA = IB/AT$. Net income scaled by total assets. Updated annually.

- **Sales Growth (SG)**:

  Follows Lakonishok et al. [1994]. $SG = SALE/SALE_{-12}$. Sales growth is the percent change in net sales over turnover (Compustat item SALE).

- **Gross Margins (GMARGINS)**:

  Follows Novy-Marx [2013]. $GMARGINS = GP/SALE$, where GP is gross profits and $SALE$ is total revenues. Rebalanced annually.

- **Asset Turnover (ATURNOVER)**:

  Follows Soliman [2008]. $ATURNOVER = SALE/AT$. Sales to total assets. Rebalanced annually.

B2 - Transaction Cost Calculation

The transaction cost calculation are based on the model proposed by Hasbrouck [2009]. This model is extended to include funding liquidity. To account for the fact that the transaction should depend on the funding liquidity, the model of Hasbrouck [2009] is redesigned by writing the transaction cost as an affine function of the funding liquidity. The model is the following:

$$m_t = m_{t-1} + \varepsilon_t \tag{2.7}$$

$$p_t = m_t + (c_0 + c_1.FR_t)q_t, \tag{2.8}$$

where m_t is the log underlying "efficient value", p_t is the log trade price, q_t is the observed trade price, q_t is a random indicator for the direction of the trade that takes the value one (minus one) if the trade took place at the ask (bid), ε_t is a random disturbance reflecting public information about the stock, and $c_t = c_0 + c_1.FL_t$ is the effective cost of trading. c_0 and c_1 are two coefficients to be determined.

By generalizing the previous equation to include a market return factor, like Hasbrouck [2009], the following equation is obtained:

$$\Delta p_t = (c_0 + c_1.FR_t)\Delta q_t + \beta_m r_{mt} + \varepsilon_t \tag{2.9}$$

This equation can be rewritten like this:

$$\Delta p_t = c_0.\Delta q_t + c_1.FR_t.\Delta q_t + \beta_m r_{mt} + \varepsilon_t \tag{2.10}$$

Like Hasbrouck [2009] mentions it, the difficulty in estimating this model of transaction cost is the fact that q_t, the random indicator for the direction of the trade, is unknown. So like Hasbrouck [2009], a bayesian approach (Gibbs sampling) is used to estimate our model. ε_t is assumed to be $iid \sim N(0, \sigma_\varepsilon^2)$. The parameters that will be estimated are c_0, c_1, β_m and σ_ε^2.

Chapter 3

Shadow Banking in the US

3.1 Introduction

Shadow banking refers to activities of financial intermediaries that are not subject to federal banking regulations. These shadow banking activities contribute to the creation of credit in the economy, but not via the traditional function of accepting deposits and giving out loans. Instead, shadow banking rely on securitization and on the creation and selling of new financial instruments and products. Examples of shadow banks, i.e. banks that rely solely on shadow activities, include hedge funds, unlisted derivatives, and other unlisted instruments. However, regulated financial institutions may also engage in shadow activities, such as credit default swaps.

In this paper we propose a novel way of measuring the shadow banking activity of firms operating in the financial sector in the United States. Because they are not subject to federal regulation, shadow banking activities are difficult to measure, and financial companies are not obligated to provide data on the extent of their involvement in shadow activities. Nonetheless, measuring the shadow banking is an important task for policy makers, especially given the major contributing role of shadow banking activities in the recent financial crisis and the Great Recession (Luttrell et al. [2012], Gorton and Metrick [2012]). The fast-paced growth in the volume of the subprime mortgages and collateralized mortgage-backed securities leading up to 2008 is the area of shadow banking admittedly responsible for the resulting financial crisis. The bursting of the associated housing bubble generated a run in the shadow banking system, which, unlike traditional banks, does not have access to a lender of last resort or federal deposit insurance. In the face of the adverse consequences of the 2008 crisis for the world economy, heightened capital requirements on traditional banks have emerged as a global trend. However, regulatory reforms remain thus far largely silent on many aspects of shadow banking activities (Adrian and Ashcraft [2012], Gorton et al. [2010]), regardless of whether they are performed by traditional or by shadow banks. In order to be able to design such reforms, regulators need to have access to a measure of shadow banking activities.

This paper proposes such a measure, using textual data from annual reports of firms operating in the US financial sector. These firms could be shadow banks, i.e. non-depository financial institutions, or traditional banks, i.e. depository financial institutions, that engage in some shadow activities, for example via Special Purpose Vehicles (SPVs) or Special Purpose Entities (SPEs). These vehicles or entities are created by a traditional parent bank so as to isolate financial risk and without being constrained by banking regulation.

The measure proposed in this paper recognizes that shadow activities can be performed by either shadow banks or by traditional banks and that, therefore, the measurement should take place at the activity level, rather than at the entity level. Thus, for example, a traditional bank may engage in some traditional activities, based on deposits, and in some shadow activities, based on securitization. The construction of the measure of shadow-banking activities relies on textual data for each firm in the US financial sector. Specifically, the 10-K and 10-Q documents are collected from the database of the Securities and Exchange Commission - Electronic Data Gathering, Analysis, and Retrieval system (SEC-EDGAR), for the period 1994-2019. A firm-level 10-K is an annual financial report that provides audited financial statements, a discussion of risk factors for company operations, and the management's analysis of prior fiscal year performance. The 10-Q form is similar to the 10-K, although less detailed, and it is issued each quarter, except for the last quarter of the year, as information for the last quarter is included in the 10-K.

First, a dictionary is constructed, consisting of words, single or composite, widely accepted by regulators and researchers as related to shadow-banking activities. Examples include terms such as "asset-backed security", "mortgage-backed security", "repos", "reverse repos", "credit default swaps", "commercial paper" etc. Second, the importance of this shadow-banking activity dictionary is calculated, for each firm, each period, and each financial document. The importance of the shadow-banking activity dictionary in a document is based on the term-frequency inverse-document frequency (TF-IDF) score for each dictionary word appearing in the document. This score increases in the TF, which measures the number of times the dictionary word appears in a document, divided by the number of words in the document. The score decreases in the IDF, which measures the number of documents that contain the word, in the corpus of documents. The corpus here refers to all the financial documents of all firms in a given year or quarter. An index of traditional banking activity is constructed along the same lines, where the dictionary contains words such as “savings account”, “checking account”, “liquidity coverage ratio” and “deposit insurance”. The idea here is that, even if numerical data about activities are not directly provided, a firm’s reports will nonetheless reflect, via the frequency and intensity of the usage of certain words, the extent of the firm’s involvement in the corresponding type of activities.

Consistent with intuition, the results show that shadow-banking activities are more intense in “Non depository Institutions”, followed by “Depository Institutions” and “Insurance carrier”. In other words, shadow firms engage in more shadow activities than traditional banks or insurance firms. Furthermore, the aggregate index of shadow activities increases continuously since 1994, reaching a maximum in 2008. This is followed by a decline, smaller in magnitude than the preceding increase, until 2012. This decline is likely related to enhanced financial regulations and reduced appetite for new financial instruments in the aftermath of the crisis. However, in after 2012, the trend reverses again, and the shadow-banking index continues to increase to this day. It should be noted that the post-2008 decline is mostly focused in the group of “Depository Institutions”. By contrast, the shadow index for “Non depository Institutions” did not decline much and it remained at higher levels than for other types of firms during the 2008-2012 period.

The measure of shadow banking activity is then validated via a variety of exercises. First, the measure is linked to variables reflecting securitization activity, such as the total real estate loans owned and securitized by finance companies, and the total financial assets of Money Market Mutual Funds (MMMFs). Specifically, the aggregate shadow-banking activity index co-moves positively with those two securitization variables. Second, the measure is linked to monetary policy and to funding liquidity. This analysis follows the evidence provided in Nelson et al. [2018], Xiao [2020], and Fontaine and Garcia [2012], who find that the money creation of shadow banks increases during periods of contractionary monetary policy by the central bank and during periods of tight funding conditions. Indeed, the results show that, when the Fed raises its fund rate, the shadow-banking activity index increases, whereas the traditional-banking activity index decreases. Furthermore, when funding liquidity gets tighter, the shadow-banking activity index increases. These results confirm that the shadow-banking activity index proposed in this paper does in fact capture the important aspects of the financial system, with respect to the interaction of traditional with shadow activities, and with the interaction of the financial firms with the central bank of the US.

Finally, the shadow-banking activity index has predictive power for loan delinquency rates. This analysis relies on evidence that households' loan delinquency rates are determined not only by economic factors, such as unemployment (Hendershott and Schultz [1993], Deng et al. [2000], and Livshits et al. [2007]), but also by the activities of lenders, especially those pertaining to securitization (Begley and Purnanandam [2017] or Jiang et al. [2014]). The analysis focuses on the 100 biggest banks in the US for the period 1994-2018. The results show that an increase in the shadow-banking activity index is associated with an increase in the loan delinquency rate for these banks, whereas an increase in traditional-banking activity index is associated with a decline in their loan delinquency rate.

Overviews of the shadow banking system are provided by Pozsar [2008], Adrian and Shin [2009] and Luttrell et al. [2012]. Shadow-banking activity measurement has been addressed in some recent papers. For example, Pozsar et al. [2010] use aggregate liabilities recorded in the Flow of Funds data relating to securitization via mortgage-backed securities (MBS), asset-backed securities (ABS), activities of government-sponsored-enterprises (GSEs), repos, commercial paper, and Money Market Mutual Funds (MMMFs). The latter three are short-term money market transactions that are not backstopped by deposit insurance. However, the accuracy of this measure is unknown, given that many of the securitized assets are held off balance sheets of traditional institutions, through backup liquidity and credit derivative or reinsurance contracts. The Financial Stability Board (FSB), Board [2013], has developed an entity-based measure of shadow banking, composed of insurance corporations, pension funds, other financial intermediaries (OFIs) and financial auxiliaries. One important potential problem with this measure is the inclusion of non-shadow banking activities by the OFIs. Methodologically, this paper is related to that of Hanley and Hoberg [2019]. Hanley and Hoberg [2019], who use computational linguistics on the 10-Ks of financial firms to develop a methodology that can detect emerging risks in the financial sector, such as risks related to real estate, prepayment, and commercial paper. Contextually, in its focus on shadow banking, this paper is also related to the work of Buchak et al. [2018], which examines potential reasons for the increase importance of shadow banking in the US housing market.

The rest of the paper is structured as follows. Section 3.2 describes shadow banking system in US and shows how it differs from traditional banking. Section 3.3 describes the data we use for our paper and presents some descriptive statistics of our dataset. In Section 3.4, we describe the methodology of our paper. We describe how we select words and phrases that compose our dictionary of shadow banking activity and how we use this dictionary to construct our shadow banking index. Section 3.5 presents the results of the shadow banking index, how it evolves from 1994 to 2019, and how it differ from one group of firm to another. In this section we also validate our shadow banking measure by linking it to variables related to the securitization activity and also by using it to verify links of shadow banking with monetary policy and funding liquidity. Section 3.6 presents the results of the use of our shadow banking index as determinant of delinquency rate. In Section 3.7 we conclude.

3.2 The US Banking System

This section describes the functions of the US banking system, especially as they pertain to the distinction between traditional banks and shadow banks.

3.2.1 Traditional Banking

In the traditional banking system, banks conduct credit intermediation by accepting deposits, which are then used for the provision of loans. This process involves three main activities, namely maturity transformation, liquidity transformation, and credit transformation. Maturity transformation refers to the fact that banks borrow on shorter time frames but lend money on longer time frames. Specifically, clients can claim their deposits at any time, whereas the bank can not claim money from borrower at anytime, as it has to respect loan deadlines. Liquidity transformation refers to the fact that a bank's assets, i.e. its loans, are less liquid than its liabilities, i.e. its deposits. Specifically, loans lave longer maturities and they face risk of borrower default. Credit transformation happens when banks invest in securities, i.e. loans, that have a lower credit rating than the bank's funding instruments, i.e. the deposits. As a result, the interest rates paid by the bank on its deposits is lower than that received from its loans.

A bank run can happen if all depositors claim their money at the same time. To prevent such runs, banks rely on the Federal Deposit Insurance Corporation (FDIC) to insure deposits and on the US central bank (Federal Reserve or Fed) as the lender of last resort. Furthermore, to ensure the health of the financial system, the federal government imposes regulations related to liquidity and capital requirements, in accordance with the Basel III guidelines.

3.2.2 Shadow Banking

The shadow banking system or shadow financial system is a network of financial institutions comprised of non-depository banks like investment banks, structured investment vehicles (SIVs), conduits, hedge funds, non-bank financial institutions and money market funds. Unlike traditional banks, shadow banks do not accept deposits, and are therefore not subject to most regulatory limits and laws imposed on the traditional banking system. They also do not have access to backstops like the FDIC or the Fed.[1]

The credit intermediation offered by shadow banks does not rely on a deposit-loan system, but rather on a complicated chain of securities and transactions. In particular, credit intermediation occurs via securitization supported by wholesale funding and obtained by the issuance of commercial papers (CP), repurchase agreements (repos), or other debt and structured credit instruments. These instruments are then sold to Money Market Mutual Funds (MMMFs), bond funds, and other entities. Since 2012, shadow banks now account for the majority of new lending in the US mortgage market.

Figure3.1 depicts the most common types of credit intermediation: a) non-intermediated, direct lending, b) intermediated lending through traditional banking, and c) intermediated lending via shadow banking.

Figure3.1 here.

3.3 Data

Our dataset is a set of forms 10-K and forms 10-Q of firms operating in the financial sector. These files are downloaded on the website of the U.S. Securities and Exchange Commission (SEC). A 10-K is an annual report that provides audited annual financial statements, a discussion of material risk factors for the company and its business, and a management's discussion and analysis of the company's results of operations for the prior fiscal year. The

[1] Special purpose vehicles (SPVs) may benefit from some liquidity requirements put in place by their parent bank in times of economic turmoil.

SEC requires one 10-K per year from each firm. The 10-Q is similar to the 10-K, however the information is generally less detailed. The 10-Q are issued each quarter, except for the last quarter of the year since information for the last quarter is included in the 10-K.

These forms are collected by EDGAR, the Electronic Data Gathering, Analysis, and Retrieval. On the SEC database a company is identified by its Central Index Key (CIK). The CIK is the unique numerical identifier assigned by the EDGAR system to filers when they sign up to make filings to the SEC. CIK numbers remain unique to the filer, they are not recycled. For our study, we use CIK of companies in the financial sector. Those companies have SIC (Standard Industrial Classification) codes between 6000 and 6999).

We have in our database $53,658$ 10-Ks and $147,898$ 10-Qs collected for $15,180$ firms. Among these $15,180$ firms, 1769 are "Depository institutions", 9787 are "Non depository institutions", 374 are "Security & Commodity Brokers", 967 are "Insurance Carriers", 35 are "Insurance Agents, Brokers, & Service", 843 are "Real Estate" and 1405 are "Holding & Other Investment Offices". New York, California, Illinois and Texas are the states with the most firms, with respectively $3,806$ firms, 2485 firms, 711 firms and 631 firms. All these informations are summarized in Tables 3.1 and 3.2.

Table 3.1 and Table 3.2 here.

The macroeconomics variables come from different sources. The state-level household debt comes from the data from the Federal Reserve[2]. The unemployment rate and Real GDP Growth rate come from the Iowa Community Indicators Program[3]. Inflation has been obtained from the US Inflation Calculator website[4]. The Fed fund rates, the Total Real Estate Loans Owned and Securitized by Finance Companies, the Money Market funds total financial assets and the TED spread are downloaded from the Federal Reserve Economic Data (FRED) of the Federal Reserve Bank of St.Louis[5].

3.4 Methodology

To measure a company's shadow banking activity, we have to track and measure in the report of this company the importance of words, phrases or abbreviations that are related to

[2] https://www.federalreserve.gov/releases/z1/dataviz/household_debt/state/map/#year:2018
[3] https://www.icip.iastate.edu/tables/employment/unemployment-states
[4] https://www.usinflationcalculator.com/inflation/historical-inflation-rates/
[5] https://fred.stlouisfed.org/

shadow banking activity. So the first thing to do is to have a dictionary of words or phrases related to shadow banking activity. This paper employs a linguistic dictionary approach. The dictionary-based approach, also referred to as the "bag-of-words" method, uses an algorithm that reads a text and finds a list of words contained in a predefined dictionary. The algorithm then gives a weight to each word in the dictionary it computes a score of the importance of the dictionary in the text. One can employ proportional weighting, which treats every word in the list as equally important, or "term frequency-inverse document frequency", which will be employed and described here.

3.4.1 Shadow Banking Word List

The paper constructs two linguistic financial dictionaries. The first includes terminology that has been used by regulators, such as the FSB and the Basel III, and by researchers to describe shadow type financial activities. The second includes terminology that has been used to describe traditional financial or banking activities. Examples of terms in the first dictionary include "asset-backed security", "mortgage-backed security", "repos", "reverse repos", "credit default swaps", "commercial paper" etc. Examples of those in the second dictionary include 'savings account", "checking account", "liquidity coverage ratio", "deposit insurance" etc. The dictionaries are presented in Appendix 3.7 and in Appendix 3.7, respectively. The next section addresses the measurement of the importance of each dictionary word in a financial document.

3.4.2 TF-IDF Weighting Scheme

In textual analysis, the most common way to measure the importance of a word in a document is through the *tf-idf* methodology, i.e. the "Term Frequency - Inverse Document Frequency" approach. This is a way to score the importance of words (or "terms") in a document based on how frequently they appear across multiple documents (the corpus). The tf captures word frequency and normalization, and the *idf* adjusts for impact across the entire corpus. Denote the corpus as D, a word or term as t, a document as $d \in D$, and the number of documents in the corpus as N. Then, the *tf-idf* value or score is given by:

$$tfidf(t, d, D) = tf(t, d).idf(t, D)$$

$$tf(t,d) = \text{frequency of } t \text{ in } d$$
$$idf(t,D) = \log \frac{N}{|\{d \in D : t \in d\}|},$$

One problem with the traditional $tf - idf$ is that it tokenizes or breaks apart any compound nouns. For example *asset backed security* will not be recognized as a compound noun, but will instead be broken up into three separate nouns, namely *asset*, *backed*, and *security*. To address this problem we modify the formula of the $tf - idf$ as follows:

$$tfidf(t,d,D) = tf(t,d).idf(t,D)$$

$$tf(t,d) = (\text{frequency of } t \text{ in } d) \times \frac{\text{number of characters in } t}{\text{number of characters in } d}$$
$$idf(t,D) = \log \frac{N}{|\{d \in D : t \in d\}|},$$

Then, the $tfidf$ of a dictionary $dict$, in a document $d \in D$, and where t is a word or group of words in the dictionary $dict$, is:

$$tfidf(dict,d,D) = \sum_{t \in dict} tfidf(t,d,D),$$

For each type of document, namely 10-Ks and 10-Qs, a separate corpus is constructed, pooling all documents for each year. Then, for each document in the corpus, the importance of the shadow banking dictionary is measured. The shadow banking index computed with the $tf - idf$ is labelled SBI_1. The results are robust to a computation that uses only proportional weights, i.e. only the TF part of the score. This index is labeled SBI_2, and is described as follows:

$$tf(t,d) = (\text{frequency of } t \text{ in } d) \times \frac{\text{number of characters in } t}{\text{number of characters in } d},$$

and therefore the importance of a dictionary $dict$ in a document d is given by:

$$tf(dict,d) = \sum_{t \in dict} tf(t,d,D).$$

3.5 Shadow-Banking Activity Index

This section presents descriptives statistics for the shadow-banking activity index derived from the linguistic analysis above. It then provides some important validation exercises for this index.

3.5.1 Descriptive Analysis

The value of the shadow-banking index is referenced with respect to the benchmark value of 2008. This benchmark is the average shadow-banking index in 2008, across all financial institutions. All the results in this paper about the shadow banking index, whether SBI_1 or SBI_2, will therefore be interpreted relatively to the average shadow banking index of the year 2008.

Figure 3.2 shows the average shadow-banking index for groups of firms defined by the SIC codes. As can be seen, shadow banking activity is more intense in the group of "Non depository Institutions". This is consistent with the definition of shadow banks as non-depository institutions. Second in intensity of shadow activities is the group of "Depository Institutions", namely the traditional banks. This means that traditional banks do in fact engage in shadow activities, and that, therefore, regulators should be focusing on the activity level, rather than the entity level. The third highest shadow banking activity is in the group of "Insurance carriers".[6]

Figure 3.2 here.

Figure 3.3 shows the evolution of the average shadow banking index from 1994 to 2019. From 1994 to 2008, the shadow banking activity is increasing and it reaches a maximum in 2008. Between 2008 and 2012, the index drops, reflecting a reluctance from supply- and demand-side to invest in shadow instruments, as well as new financial regulations put in place. However, this decline is relatively short-lived, as after 2012 the shadow banking index starts increasing again.

Figure 3.3 here.

[6]Shadow banking activities are also conducted by insurance companies, as shown in (Adrian and Ashcraft [2016]).

Figure 3.5 shows the evolution of the average shadow banking index from 1994 to 2019 for each group of firms. Before 2008, the shadow banking index was highest for the "Depository Institutions", "Non depository Institutions" and "Insurance Carriers". After 2008, the shadow banking index declines for most financial companies, except for the "Non depository Institutions". For those firms, the index dropped a bit but it stayed at a higher level, compared to the others. This is consistent with the implementation of Basel III, which placed new controls on firms that could be regulated as "Depository Institutions", but not on firms like "Non depository Institutions". This finding is consistent with the with the evidence in Adrian and Ashcraft [2012] and Gorton et al. [2010].

Figure 3.5 here.

Figure 3.4 presents the evolution of the average shadow banking index from 1994 to 2019 for the four states that account for most of the financial documents in the data, namely New York, California, Illinois and Texas. New York clearly has the highest shadow banking activity, consistent with its role as a global financial centner.

Figure 3.4 here.

3.5.2 Co-movement with Securitisation

Because shadow banks give contractual debts such as residential mortgages, commercial mortgages, auto loans or credit card, and pooled these debts (securitization), and then sell them through wholesale funding to structures like MMMFs, the shadow-banking activity index should be strongly positively related to securitization. This section then examines how the average shadow banking index co-moves with some variables that are well-known to be related to securitization.

Figure 3.6 presents the evolution of the average shadow banking index along with the Total Real Estate Loans Owned and Securitized by Finance Companies. As can be seen, the shadow banking index follows the movement of Real Estate Loans with a small lag. Figure 3.7 presents the evolution of the average shadow banking index along with the total financial assets of MMMFs. Clearly, there is a positive co-movement between the shadow bank activity index and the total financial assets of MMFs. This analysis confirms that the shadow banking index does in fact capture securitization activities.

Figure 3.6 and Figure 3.7 here.

3.5.3 Monetary Policy and Funding Liquidity

Using vector autoregressive models for the US, Nelson et al. [2018] find that a contractionary monetary policy shock has a persistent negative effect on the asset growth of commercial banks, but a persistent positive effect on the asset growth of shadow banks. Furthermore, Xiao [2020] find that shadow-bank money creation significantly expands during monetary tightening, and that this "shadow banking channel" offsets reductions in commercial bank deposits and dampens the impact of monetary policy on traditional banks. A qualitatively similar result for funding liquidity is presented in Fontaine and Garcia [2012].

This section seeks to validate these findings via the use of the average shadow banking index for "Depository Institutions", termed SBI_1, and for "Non depository Institutions", termed SBI_2. The shadow index for each bank type is the dependent variable. The regression equation for each shadow index is:

$$SBI_t = \alpha + \beta_{fed} \cdot Fed\_Rate_t + \beta_{gdp\_growth} \cdot Growth + \beta_{infl} \cdot Inflation + \beta_{Ted\_Spread} \cdot Ted\_Spread + u_t \tag{3.1}$$

We also run another regression in which the dependant variable is the traditional banking index. The traditional banking index has been constructed similarly to the shadow banking index, the only difference being the dictionary of words and phrases. The list of words in this dictionary is presented in Appendix 3.7. The traditional banking index computed with the *tf-idf* is labelled TBI_1 and the shadow banking index computed with just the term frequency is labelled TBI_2.

The time period is quarterly. As in Xiao [2020], we also include two measures of regulatory tightness: A dummy variable for the Gramm-Leach-Bliley Act, which loosened financial regulations in 1999; and a dummy variable for the Dodd-Frank Act, which tightened financial regulations in 2010. Here, β_{fed} is expected to be positive, if indeed the shadow index increases when monetary policy contracts. The results of Equation 3.1 are presented in Table 3.3. As shown in columns (1) and (2), the coefficient of the Fed funds rate is positive. This means that, when the Fed raises its rate, then shadow banking activity increases. The opposite is observed with the traditional banking index, in columns (3) and (4). In other words, when the Fed raises its rate, then traditional banking activity decreases. The coefficient of the TED spread in columns (1) and (2) is also positive. This suggests that, when funding liquidity gets tighter, then shadow banking activity increases.

Table 3.3 here.

Overall, the results in this section confirm the validity of the shadow-banking index as capturing securitization activities, with intensity that depends on monetary and financial policies, consistent with previous evidence in the literature.

3.6 Shadow-Banking Activity Index and Delinquency Rates

A number of studies have examined the determinants of delinquency on household loans. One important conclusion is that macroeconomic factors, especially unemployment, are a significant determinant of household delinquency and bankruptcy decisions.[7]. A second conclusion is that delinquency can also be due to lenders' activities of the lender, specifically the link between securitization and loan performance.[8].

This section investigates the determinants of the loan delinquency rates of the 100 biggest US banks. The dependent variables include important macroeconomic factors, as well as the traditional- and shadow- bank activity indexes constructed in this paper. Let DR_t be the delinquency rate in quarter t. The regression equation for the 100 biggest US banks is:

$$DR_t = \alpha + \beta_1.SBI_t + \beta_2.TBI_t + \beta_3.Unemp_t + \beta_4.Infl_t + \beta_5.Growth_t + \beta_6.Fedfund_t + u_t, \quad (3.2)$$

where SBI_t is the average shadow banking index of these banks, TBI_t is the traditional banking index of these banks, $Unemp_t$ is the unemployment rate, $Infl_t$ is the inflation rate, $Growth_t$ is the growth rate of real GDP, and $Fedfund_t$ is the average Fed funds rate.

Table 3.5 presents the results of Equation 3.2, where the indexes use the TF-IDF as the weighting scheme. Columns (1) and (2) use the delinquency rate on all real estate loans. Columns (3) and (4) use the delinquency rate on residential real estate loans. Column (5) and (6) use the delinquency rate on credit cards. Columns (7) and (8) use the delinquency rate on all loans. Table 3.6 repeats the analysis with proportional weights.

Table 3.5 and Table 3.6 here.

As can be seen in both Table 3.5 and Table 3.6, the coefficients of the shadow-banking activity index are positive while those of the traditional banking activities are negative. This suggests

[7]See Hendershott and Schultz [1993], Deng et al. [2000], and Livshits et al. [2007])

[8]See (Begley and Purnanandam [2017] and Jiang et al. [2014])

that an increase in the shadow banking activity of the 100 biggest US banks is associated with an increase in the delinquency rates on the loans given by these banks. The opposite is true for the traditional banking coefficients. Here, an increase in traditional banking activity is associated with a decrease in the loan delinquency rates in the 100 biggest US banks. Furthermore, the coefficient of the Fed funds rate are positive, indicating that an increase in policy rates is associated with a decrease in loan default rates in the 100 biggest banks, likely because of an associated increase in commercial bank rates.

3.7 Conclusion

In this paper, we propose a new way of measuring shadow banking activities through textual analysis of reports filed by financial companies in the US. The results show that all banks may engage in shadow activities, defined as credit intermediation activities relying heavily on securitization. In turn, this means that the appropriate definition of shadow banking is at the activity level, as opposed to the entity level, with potential associated implications for measurement, financial regulation, and monetary policy. Overall, since 1994, shadow banking activities have been increasing at a fast pace for all financial institutions in the US, but especially for those designated as "Non depository Institutions". Even for "Depository Institutions" and "Real estate" companies, the trend has been an increasing one, despite a slowdown during the period 2008-2012, in the aftermath of the financial crisis. Furthermore, the shadow index proposed co-moves strongly and positively with accepted measures of securitization, such as the total real estate loans owned and securitized by finance companies, and the total financial assets of MMMFs. In addition, contractionary monetary policy and/or tighter funding liquidity are associated with an increase in the shadow activity index, confirming hypotheses that have been proposed in the literature. What is more, for the sample of the 100 biggest US banks, an increase in the shadow banking index is associated with an increase in households' loan delinquency rates, whereas an increase in the traditional banking index is associated with a decline in those rates.

Overall, two main results emerge. First, the intensity of shadow-banking activities can be influenced by monetary and by macro-prudential policies. Second, the intensity of shadow-banking activities has predictive power for households' loan delinquency rates. This could be because of the lack of regulation on the shadow instruments offered or because of lower credit-worthiness of the household willing to trade in shadow instruments. In both cases, the results for the macroeconomy could be adverse, thereby offering a role for regulatory policy in relation to household welfare.

Tables and Figures of Chapter 3

Table 3.1: **Number of reports by SIC code**

SIC code	Number of 10-Ks	Number of 10-Qs	Number of firms
60 Depository Institutions	14,574	43,722	1,769
61 Non depository Institutions	17,851	44,628	9,787
62 Security & Commodity Brokers	2,946	8,838	374
63 Insurance Carriers	3,970	11,908	967
64 Insurance Agents, Brokers, & Service	319	942	35
65 Real Estate	5,750	14,950	843
67 Holding & Other Investment Offices	8,248	22,910	1,405
Total	53,658	147,898	15,180

Table 3.2: **Number of reports by State**

State	**Number of 10-Ks**	**Number of firms**	**Deposit inst.**
New York	10,807	3,806	120
California	6,817	2,485	158
Illinois	3,150	711	97
Texas	2,740	631	48
Pennsylvania	2,324	294	148
Maryland	2,054	853	50
Florida	1,778	378	67
North Carolina	1,803	621	63
Massachusetts	1,854	385	64
New Jersey	1,583	383	89
Virginia	1,629	491	81
Ohio	1,494	175	89
Connecticut	1,266	368	12
Delaware	1,059	538	6
Georgia	1,176	181	72
Michigan	1,171	241	57
Others	10,953	2639	548
Total	53,658	15,180	1,769

Table 3.3: **Monetary policy, Shadow Banking Index and Traditional Banking Index**

This table presents time-series regressions of shadow banking activity index on rate and Ted spread in columns **(1)** and **(2)** and regressions of traditional banking activity index on Fed fund rates and Ted spread in columns **(3)** and **(4)**. Many control variables such as the real GDP growth rate, Inflation rate, a dummy variable for the Gramm-Leach-Bliley Act which loosens financial regulation, a dummy variable for the Dodd-Frank Act which tightens financial regulation are also included in the regression. A time trend is also included in the regressions. The data frequency is quarterly. The sample period is from 1994 to 2019. The Newey-West standard errors with 6 lags are presented in brackets. $*p > 0.1$; $**p > 0.05$; $***p > 0.01$.

Covariates	**(1)**	**(2)**	**(3)**	**(4)**
Fed Fund rates	2.646 **	2.976 ***	-6.617 **	-4.009 **
	(1.058)	(1.082)	(3.233)	(1.956)
Ted spread	11.776 ***	21.255 ***	-9.610	-5.315
	(2.384)	(4.270)	(9.720)	(6.615)
GDP growth	-0.010	1.172	-1.291	-6.010 **
	(0.686)	(1.161)	(3.775)	(2.507)
Inflation	3.022 ***	1.657	7.997 *	6.132 **
	(0.889)	(2.217)	(4.248)	(2.930)
Gramm-Leach-Bliley	1.293 ***	-3.652	14.063	3.965
	(4.687)	(4.882)	(11.950)	(7.762)
Dodd-Frank	-30.089 ***	-33.289 ***	-31.206	10.142
	(9.689)	(10.571)	(32.275)	(23.635)
Trend	0.573 ***	1.242 ***	-1.136	-1.556 ***
	(0.198)	(0.238)	(0.682)	(0.473)
Cst	30.330	-22.059	160.712 ***	182.247 ***
	(7.137)	(12.557)	(31.853)	(23.980)
Adj R-squared	66.68%	70.19%	63.97%	70.30%
N	104	104	104	104

Table 3.4: **100 Biggest bank in US as of December 2019**

Rank	Bank name	Headquarters	Total assets (billions of US$)
1	JPMorgan Chase	New York City	2,687
2	Bank of America	Charlotte, North Carolina	2,434
3	Citigroup	New York City	1,951
4	Wells Fargo	San Francisco, California	1,927
5	Goldman Sachs	New York City	992
6	Morgan Stanley	New York City	895
7	U.S. Bancorp	Minneapolis,Minnesota	495
8	Truist Financial	Charlotte, North Carolina	473
9	PNC Financial Services	Pittsburgh,Pennsylvania	410
10	TD Bank, N.A.	Cherry Hill, New Jersey	408
11	Capital One	McLean, Virginia	390
12	The Bank of New York Mellon	New York City	381
13	TIAA	New York City	315
14	Charles Schwab Corporation	San Francisco, California	294
15	HSBC Bank USA	New York City	249
16	State Street Corporation	Boston,Massachusetts	245
17	American Express	New York City	198
18	Ally Financial	Detroit,Michigan	180
19	State Farm	Bloomington, Illinois	178
20	USAA	San Antonio, Texas	173
21	BMO Harris Bank	Chicago,Illinois	172
22	MUFG Union Bank	New York City	170
23	Fifth Third Bank	Cincinnati,Ohio	169
24	Citizens Financial Group	Providence, Rhode Island	166
25	Ameriprise	Minneapolis	151
26	Santander Bank	Boston,Massachusetts	149
27	Barclays	New York City	149
28	KeyCorp	Cleveland,Ohio	145
29	RBC Bank	New York City	139
30	UBS	New York City	139
31	Northern Trust	Chicago,Illinois	136
32	Regions Financial Corporation	Birmingham, Alabama	126
33	BNP Paribas	New York City	125
34	M&T Bank	Buffalo, New York	119
	First Republic Bank	San Francisco	116
35	Credit Suisse	New York City	114
36	Discover Financial	Riverwoods, Illinois	113
37	Deutsche Bank	New York City	109
38	Huntington Bancshares	Columbus, Ohio	109
39	Synchrony Financial	Stamford, Connecticut	104
40	BBVA USA	Birmingham,Alabama	93
41	Comerica	Dallas,Texas	73
42	Silicon Valley Bank	Santa Clara, California	71
	Zions Bancorporation	Salt Lake City,Utah	69
43	E-Trade	New York City	61
44	People's United Financial	Bridgeport, Connecticut	58
45	New York Community Bank	Westbury, New York	53

46	Popular, Inc.	San Juan, Puerto Rico	52
47	CIT Group	New York City	50
48	Mutual of Omaha	Omaha, Nebraska	50
49	Synovus	Columbus, Georgia	48
50	CIBC Bank USA	Chicago, Illinois	42
51	TCF Financial	Detroit, Michigan	47
52	East West Bank	Pasadena, California	46
53	Mizuho Financial Group	New York City	43
54	First Horizon National Corporation	Memphis, Tennessee	43
55	BOK Financial Corporation	Tulsa, Oklahoma	42
56	Raymond James Financial	St. Petersburg, Florida	40
57	First Citizens BancShares	Raleigh, North Carolina	39
58	John DeereCapital Corporation	Reno, Nevada	39
59	Valley National Bank	Wayne, New Jersey	37
60	Wintrust Financial	Rosemont, Illinois	36
61	FNB Corporation	Pittsburgh, Pennsylvania	34
62	Frost Bank	San Antonio, Texas	34
63	BankUnited	Miami Lakes, Florida	32
64	Texas Capital Bank	Dallas, Texas	32
65	Associated Banc-Corp	Green Bay, Wisconsin	32
66	Prosperity Bancshares	Houston, Texas	32
67	IberiaBank	Lafayette, Louisiana	31
68	Sterling Bancorp	Montebello, New York	30
69	Hancock Whitney	Gulfport, Mississippi	30
70	Webster Bank	Waterbury, Connecticut	30
71	Umpqua Holdings Corporation	Portland, Oregon	28
72	Pinnacle Financial Partners	Nashville, Tennessee	27
73	Western Alliance Bank	Phoenix, Arizona	26
74	Investors Bank	Short Hills, New Jersey	26
75	PacWest Bancorp	Los Angeles, California	26
76	UMB Financial Corporation	Kansas City, Missouri	26
77	Commerce Bancshares	Kansas City, Missouri	26
78	Stifel	St. Louis, Missouri	24
79	Flagstar Bank	Troy, Michigan	23
80	MidFirst Bank	Oklahoma City, Oklahoma	23
81	Sumitomo Mitsui Financial Group	New York City	22
82	First National of Nebraska	Omaha, Nebraska	22
83	Macy's	Cincinnati, Ohio	22
84	Fulton Financial Corporation	Lancaster, Pennsylvania	21
85	Simmons Bank	Pine Bluff, Arkansas	21
86	Old National Bank	Evansville, Indiana	20
87	First Hawaiian Bank	Honolulu, Hawaii	20
88	FirstBank Holding Co	Lakewood, Colorado	19
89	United Bank (West Virginia)	Charleston, West Virginia	19
90	Arvest Bank	Bentonville, Arkansas	19
91	Ameris Bancorp	Atlanta, Georgia	18
92	Bank of Hawaii	Honolulu, Hawaii	18
93	Cathay Bank	Los Angeles, California	18
94	First Midwest Bank	Chicago, Illinois	17
95	Cadence Bank	Atlanta, Georgia	17
96	Atlantic Union Bank	Richmond, Virginia	17
97	Mechanics Bank	Walnut Creek, California	17

98	CenterState Bank	Winter Haven, FL	17
99	Washington Federal	Seattle, Washington	16
100	South State Bank	Columbia, South Carolina	15

Table 3.5: **Determinants of delinquency rates of the 100 biggest banks**

This table presents time-series regressions of delinquency rates on shadow banking activity index and traditional banking activity index. The shadow banking index and the traditional banking index used are the SBI_1 and TBI_1 which are the indexes computed with the *tf-idf*. Many control variables such as the unemployment rate,the real GDP growth rate,Inflation rate and Fed fund rate are also included in the regression. A time trend is also included in the regressions. The data frequency is quarterly. The sample period is from 1994 to 2019. The Newey-West standard errors with 6 lags are presented in brackets. $*p > 0.1$; $**p > 0.05$; $***p > 0.01$.

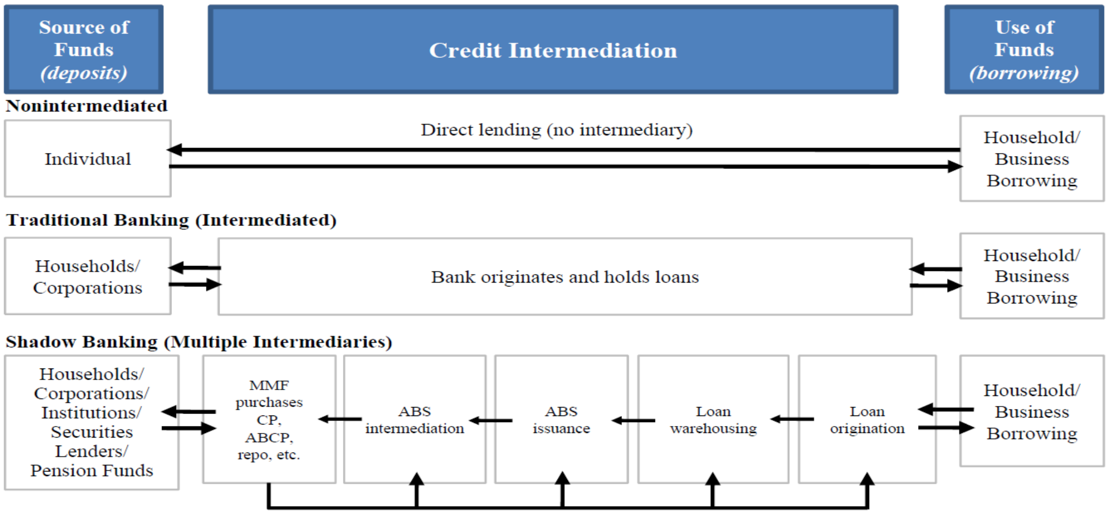
	Real estate		**R**esidential		**C**redit card		**T**otal loans	
Covariates	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
SBI_1	0.0470 ***	0.0026	0.0548 ***	0.0017	0.0279 **	0.0102 **	0.0239 **	0.0112 **
	(0.0175)	(0.0061)	(0.0185)	(0.0079)	(0.0119)	(0.0050)	(0.0113)	(0.0032)
TBI_1	-0.0293 ***	-0.0074 **	-0.0360 ***	-0.0038	-0.0108 **	-0.0071 **	-0.0114 **	-0.0101 ***
	(0.0070)	(0.0032)	(0.0089)	(0.0047)	(0.0049)	(0.0026)	(0.0049)	(0.0019)
UNEMP		1.9139 ***		2.0789 ***		0.2080		1.0378 ***
		(0.1324)		(0.02060)		(0.1408)		(0.0818)
GROWTH		0.1052		0.2173 *		-0.3090 ***		-0.0845
		(0.0821)		(0.1211)		(0.0752)		(0.0563)
INFL		-0.0612		-0.1715		-0.0660		-0.0535
		(0.1735)		(0.2280)		(0.0894)		(0.0847)
FED RATE		0.4776 **		0.5074 *		0.1889 *		0.2284 **
		(0.2130)		(0.2700)		(0.1022)		(0.0937)
TREND		0.0394 ***		0.0743 ***		-0.0163 **		0.0127 **
		(0.0107)		(0.0140)		(0.0066)		(0.0049)
Cst	3.5320 ***	-9.9641 ***	4.1514 ***	-12.6286 ***	2.6784 ***	3.3461 ***	2.4293 ***	-3.7618 ***
	(1.0502)	(1.6620)	(1.3262)	(2.1517)	(0.2961)	(1.2254)	(0.5980)	(0.8988)
Adj R-squared	12.52%	92.29%	12.91%	89.84%	17.83%	66.14%	5.99%	92.09%
N	106	106	106	106	106	106	106	106

Table 3.6: **Determinants of delinquency rates of the 100 biggest banks**

This table presents time-series regressions of delinquency rates on shadow banking activity index and traditional banking activity index. The shadow banking index and the traditional banking index used are the SBI_2 and TBI_2 which are the indexes computed with the term frequency. Many control variables such as the unemployment rate,the real GDP growth rate,Inflation rate and Fed fund rate are also included in the regression. A time trend is also included in the regressions. The data frequency is quarterly. The sample period is from 1994 to 2019. The Newey-West standard errors with 6 lags are presented in brackets. $*p > 0.1$; $**p > 0.05$; $***p > 0.01$.

	Real estate		**R**esidential		**C**redit card		**T**otal loans	
Covariates	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
SBI_2	0.0380 **	0.0116 *	0.0467 **	0.0066	0.0227 **	0.0124 **	0.0171	0.0015
	(0.0178)	(0.0068)	(0.0233)	(0.0108)	(0.0118)	(0.0054)	(0.0116)	(0.0038)
TBI_2	-0.0235 *	-0.0146 ***	-0.0358 ***	-0.0093 **	-0.0198 ***	-0.0113 ***	-0.0042	-0.0009
	(0.0123)	(0.0062)	(0.0175)	(0.0091)	(0.0062)	(0.0021)	(0.0063)	(0.0029)
UNEMP		1.9586 ***		2.1043 ***		0.1783		1.0323 ***
		(0.1243)		(0.1946)		(0.1232)		(0.0825)
GROWTH		0.0949		0.2063 *		-0.2843 ***		-0.0773
		(0.0780)		(0.1188)		(0.0676)		(0.0580)
INFL		-0.0347		-0.1469		-0.1023		-0.0572
		(0.1483)		(0.2122)		(0.0899)		(0.0774)
FED RATE		0.5037 **		0.5203 **		0.1753 *		0.2273 **
		(0.2002)		(0.2542)		(0.0938)		(0.0943)
TREND		0.0389 ***		0.0731 ***		-0.0147 ***		0.0126 **
		(0.0101)		(0.0138)		(0.0055)		(0.052)
Cst	3.3357 **	-10.4267 ***	4.2450 **	-12.7112 ***	2.2457 ***	3.1209 ***	2.1841 **	-3.9167 ***
	(1.5576)	(1.6588)	(1.9148)	(2.0690)	(0.4732)	(1.1122)	0.8437	(0.9855)
Adj R-squared	3.21%	92.55%	5.59%	90.06%	34.60%	70.24%	0.31%	90.78%
N	106	106	106	106	106	106	106	106

Figure 3.1: **Channels of Financial Intermediation (Luttrell et al. [2012])**



NOTE: MMF is money market mutual fund, CP is commercial paper, ABCP is asset-backed CP, repo is repurchase agreement, and ABS is asset-backed securities. See Glossary for definitions.

Figure 3.2: **Average Shadow banking activity (from 10-Ks) by SIC code**

SBI_1 is the shadow banking index we get when we use *tf-idf* as the importance of a word in a document and SBI_2 is the shadow banking index we get when we use the term frequency.

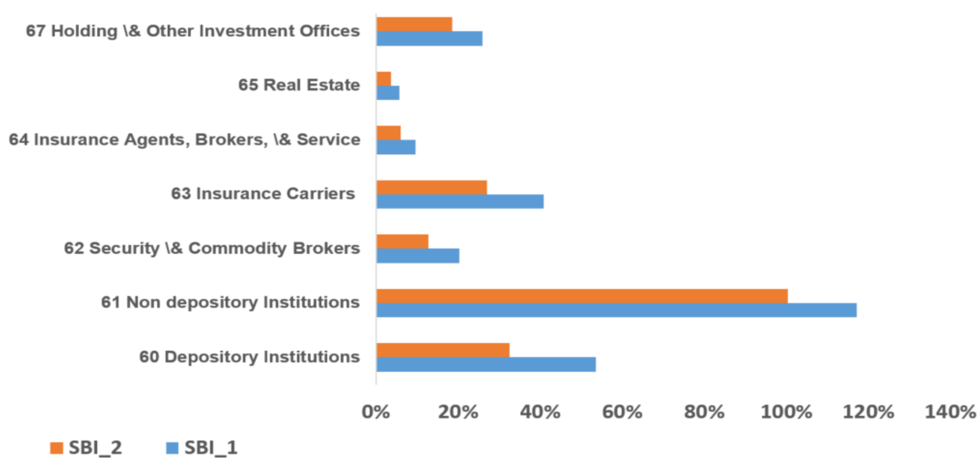


Figure 3.3: **Evolution from 1994 to 2019 of shadow banking (from 10-Ks)**

SBI_1 is the shadow banking index we get when we use *tf-idf* as the importance of a word in a document and SBI_2 is the shadow banking index we get when we use the term frequency.

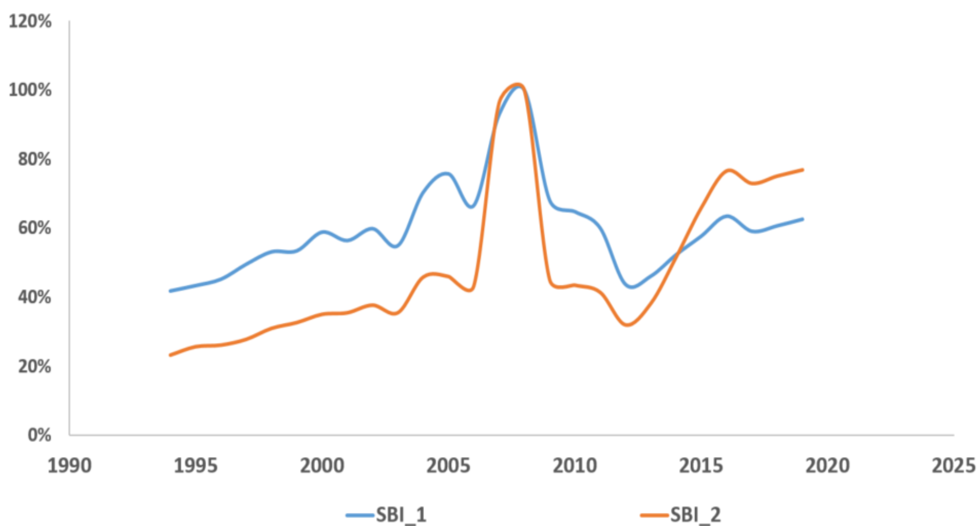
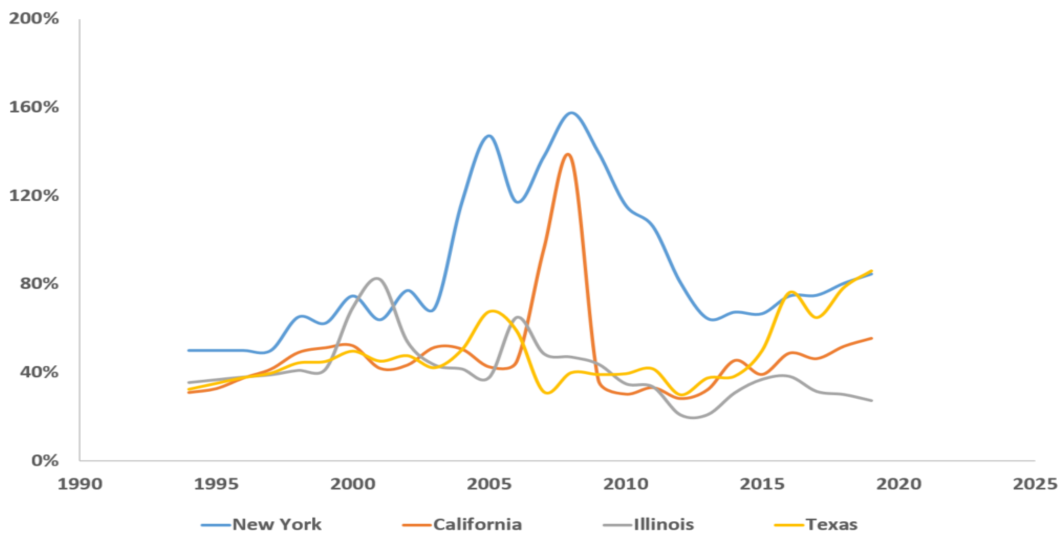


Figure 3.4: **Evolution from 1994 to 2019 of shadow banking by the state in which the firm is located (from 10-Ks)**

On the first graph, the shadow banking index presented is SBI_1, the index we get when we use *tf-idf* as the importance of a word in a document and SBI_2, the index we get when we use the term frequency is presented on the second graph.



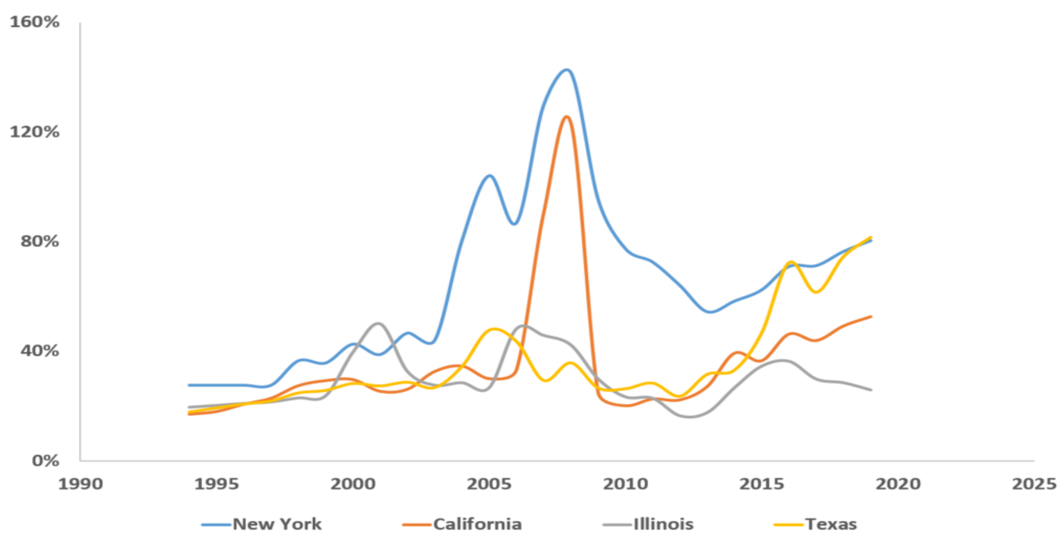
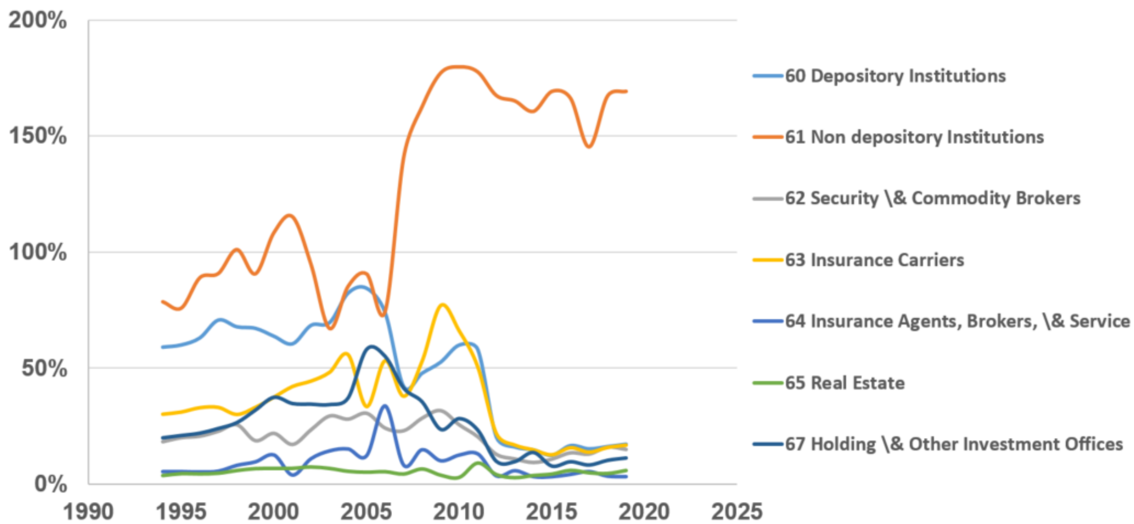


Figure 3.5: **Evolution by type of financial company from 1994 to 2019 of shadow banking (from 10-Ks)**

On the first graph, the shadow banking index presented is SBI_1, the index we get when we use *tf-idf* as the importance of a word in a document and SBI_2, the index we get when we use the term frequency is presented on the second graph.



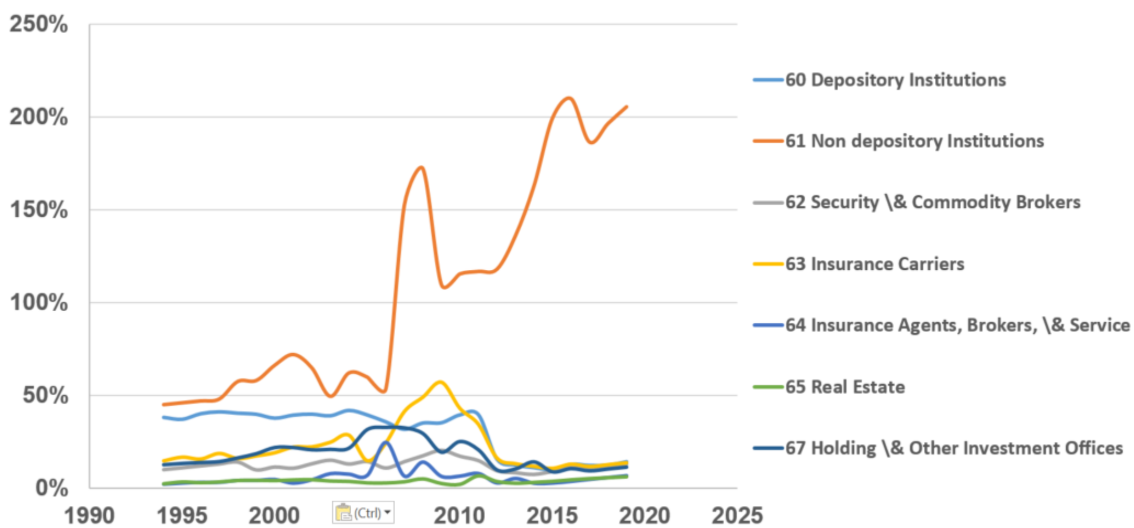


Figure 3.6: **Shadow banking and Total Real Estate Loans Owned and Securitized by Finance Companies**

On the first graph, the shadow banking index presented is SBI_1, the index we get when we use *tf-idf* as the importance of a word in a document and SBI_2, the index we get when we use the term frequency is presented on the second graph.

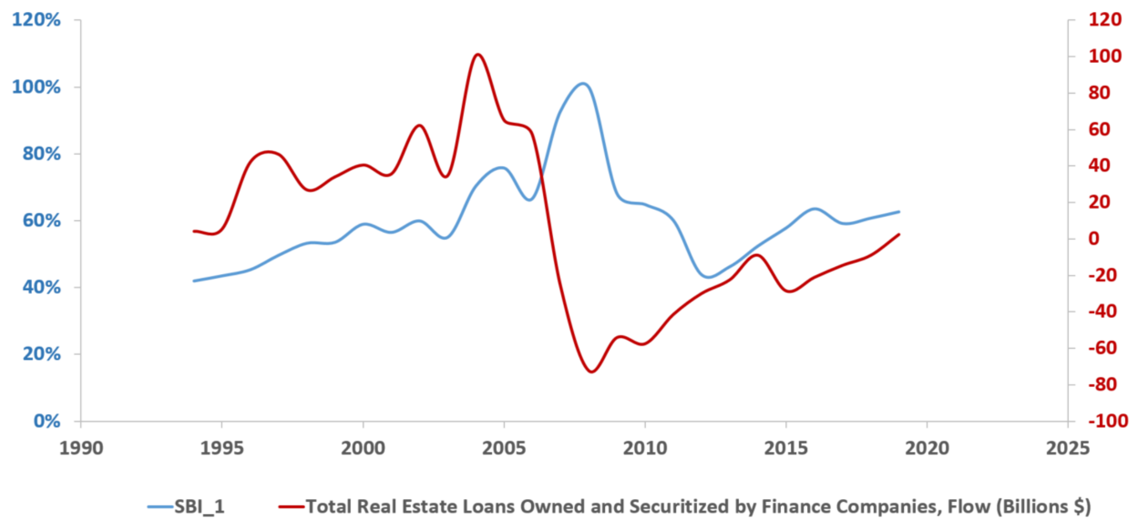
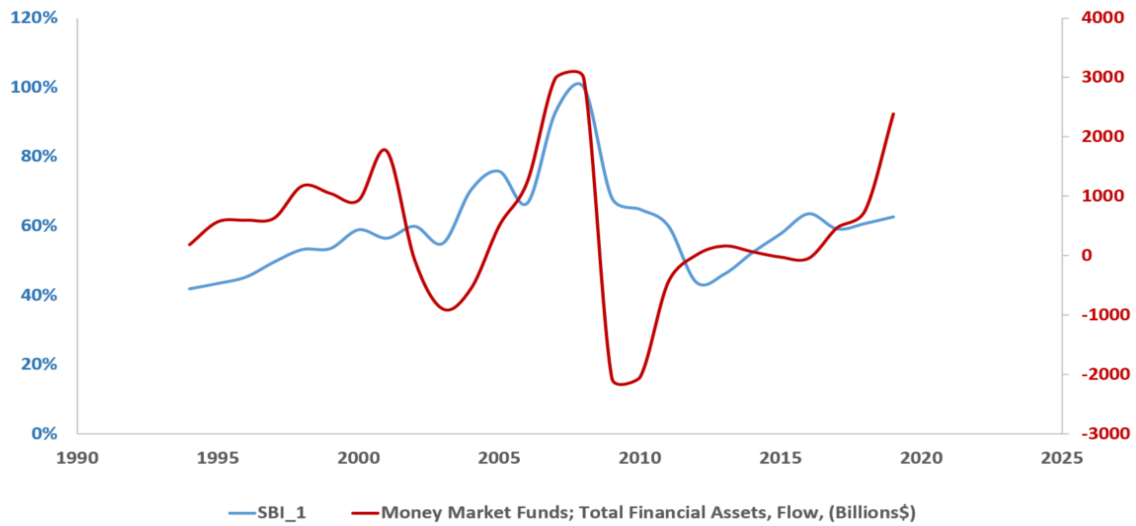
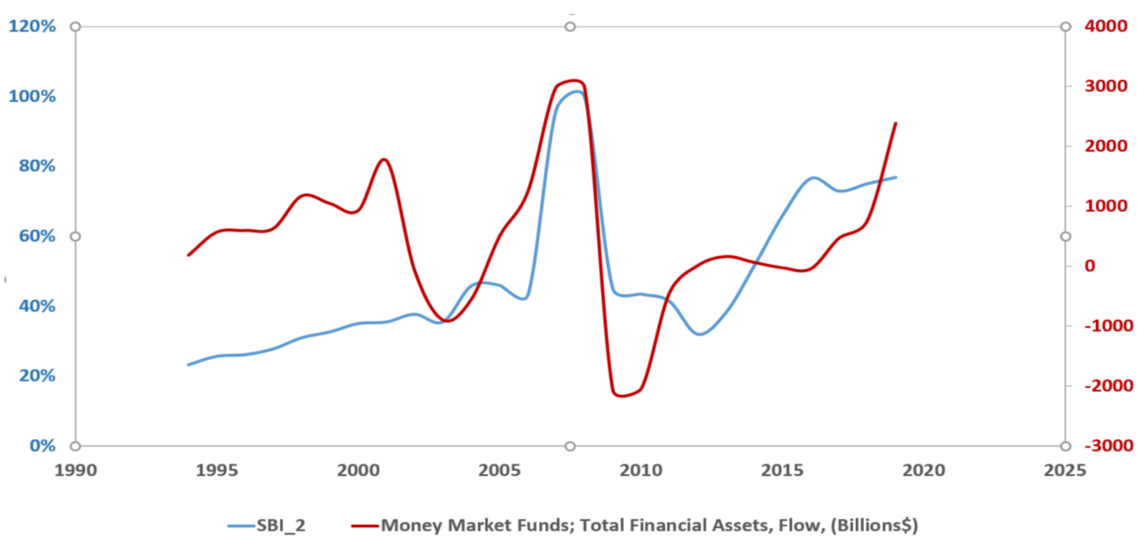


Figure 3.7: **Shadow banking and Money Market funds total financial assets**

On the first graph, the shadow banking index presented is SBI_1, the index we get when we use *tf-idf* as the importance of a word in a document and SBI_2, the index we get when we use the term frequency is presented on the second graph.





Appendices for Chapter 3 (C)

C1 - Dictionary of Shadow Banking Words

- "Securitized debt": Debt is securitized by pooling certain types of debt instruments and creating a new financial instrument from the pooled debt. The types of debt instruments used may include residential mortgages, commercial mortgages, car loans or credit card obligations. The banks receive fees for selling the new debt security.

- "Non-conforming mortgage" : A non-conforming mortgage is a term in the United States for a residential mortgage that does not conform to the loan purchasing guidelines set by the Federal National Mortgage Association /Federal Home Loan Mortgage Corporation. Mortgages which are non-conforming because they have a dollar amount over the purchasing limit set by FNMA/FHLMC are often called "jumbo" mortgages.

- "Jumbo mortgage": Another calling for the non-conforming mortgage. Also called jumbo loan.

- "Alternative A-paper": An Alternative A-paper is a type of U.S. mortgage that, for various reasons, is considered riskier than A-paper, or "prime", and less risky than "subprime," the riskiest category.

- "Alt-A mortgage": An Alt-A mortgage, short for Alternative A-paper.

- "Subprime' or "sub-prime": Subprime is a classification of borrowers with a tarnished or limited credit history. Lenders will use a credit scoring system to determine which loans a borrower may qualify for. Subprime loans carry more credit risk, and as such, will carry higher interest rates as well. Approximately 25% of mortgage originations are classified as subprime. The term subprime gets its name from the prime rate, which is the rate at which people and businesses with excellent credit history are allowed to borrow money.

- “Asset-backed security”: An asset-backed security (ABS) is a financial security collateralized by a pool of assets such as loans, leases, credit card debt, royalties or receivables.
- “ ABS ”: Abbreviation for asset-backed security.
- “ Mortgage-backed security ”: A mortgage-backed security (MBS) is a type of asset-backed security that is secured by a mortgage or collection of mortgages. This security must also be grouped in one of the top two ratings as determined by an accredited credit rating agency, and usually pays periodic payments that are similar to coupon payments. Furthermore, the mortgage must have originated from a regulated and authorized financial institution.
- “ MBS ”: Abbreviation for mortgage-backed security.
- “Residential mortgage-backed securities”: Residential mortgage-backed securities (RMBS) are a type of mortgage-backed debt obligation created from residential debt, such as mortgages, home-equity loans and subprime mortgages. A residential mortgage-backed security is comprised of a pool of mortgage loans created by banks and other financial institutions. The cash flows from each of the pooled mortgages is packaged by a special-purpose entity into classes and tranches, which then issues securities and can be purchased by investors.
- “ RMBS ”: Abbreviation for Residential mortgage-backed securities.
- “Commercial paper”: Commercial paper is an unsecured, short-term debt instrument issued by a corporation, typically for the financing of accounts payable and inventories, and meeting short-term liabilities. Maturities on commercial paper rarely range longer than 270 days. Commercial paper is usually issued at a discount from face value and reflects prevailing market interest rates.
- “Commercial mortgage-backed securities”: Commercial mortgage-backed securities (CMBS) are a type of mortgage-backed security that is secured by mortgages on commercial properties, instead of residential real estate. A CMBS can provide liquidity to real estate investors and commercial lenders. As with other types of MBS, the increased use of CMBS can be attributable to the rapid rise in real estate prices over the years.

- “ CMBS ”: Abbreviation for Commercial mortgage-backed securities.

- “ Collateralized mortgage obligations ”: Collateralized mortgage obligation (CMO) refers to a type of mortgage-backed security that contains a pool of mortgages bundled together and sold as an investment. Organized by maturity and level of risk, CMOs receive cash flows as borrowers repay the mortgages that act as collateral on these securities. In turn, CMOs distribute principal and interest payments to their investors based on predetermined rules and agreements.

- “ CMO ” or “ CMOs ”: Abbreviation for Collateralized mortgage obligations.

- “ Collateralized loan obligation ”: A collateralized loan obligation (CLO) is a single security backed by a pool of debt. Often these are corporate loans that have a low credit rating, or leveraged buyouts made by a private equity firm to take a controlling interest in an existing company. Collateralized loan obligations are similar to collateralized mortgage obligations (CMOs), except that the underlying debt is of a different type and character (companies' loans instead of mortgages).

- “ CLO ” or “ CLOs ”: Abbreviation for Collateralized loan obligations.

- “ Collateralized Bond Obligation ”: Collateralized Bond Obligation (CBO) is an investment-grade bond backed by a pool of junk bonds. Junk bonds are typically not investment grade, but because the pool includes several types of credit quality bonds together, they offer enough diversification to be ”investment grade.”

- “ CBO ” or “ CBOs ”: Abbreviation for Collateralized Bond obligations.

- “ Collateralized debt obligation ”: A collateralized debt obligation (CDO) is a structured financial product that pools together cash flow-generating assets and repackages this asset pool into discrete tranches that can be sold to investors. A collateralized debt obligation is named for the pooled assets - such as mortgages, bonds and loans - that are essentially debt obligations that serve as collateral for the CDO.

- “ CDO ” or “ CDOs ”: Abbreviation for Collateralized Debt obligations.

- “ Credit default swap ”: A credit default swap (CDS) is a financial derivative or contract that allows an investor to ”swap” or offset his or her credit risk with that of another investor. For example, if a lender is worried that a borrower is going to default on a loan, the lender could use a CDS to offset or swap that risk. To swap the risk of default, the lender buys a CDS from another investor who agrees to reimburse the lender in the case the borrower defaults. Most CDS will require an ongoing premium payment to maintain the contract, which is like an insurance policy

- " CDS ": Abbreviation for credit default swap.

- " Structured product ": A structured product is a pre-packaged investment strategy based on a single security, a basket of securities, options, indices, commodities, debt issuance or foreign currencies, and to a lesser extent, derivatives. There two types of structured products: structured deposits and structured investments.

- " Structured deposits": Structured deposits are savings accounts, offered from time to time by some banks, building societies and National Savings & Investments, where the rate of interest you get depends on how the stock market index or other measure performs. If the stock market index falls, you will usually get no interest at all.

- "Structured investments": Structured investments are commonly offered by insurance companies and banks. Your money typically buys two underlying investments, one to protect your capital and another to provide the bonus. The return you get depends on how the stock market index or other measure performs. In addition, if it performs badly or the firms providing the underlying investments fail, you might lose some or all of your original investment.

- " Structured note ": A structured note is a debt obligation that also contains an embedded derivative component that adjusts the security's risk/return profile. The return performance of a structured note will track both that of the underlying debt obligation and the derivative embedded within it. This type of note is a hybrid security that attempts to change its profile by including additional modifying structures, therefore increasing the bond's potential returns.

- "Hybrid security": A hybrid security is a single financial security that combines two or more different financial instruments. Hybrid securities, often referred to as hybrids, generally combine both debt and equity characteristics. The most common type of hybrid security is a convertible bond that has features of an ordinary bond but is heavily influenced by the price movements of the stock into which it is convertible.

- “Tranche”: Tranches are pieces of debt or securities designed to divide risk or group characteristics in order to be marketable to different investors. Each portion, or tranche, is one of several related securities offered at the same time but with varying risks, rewards and maturities to appeal to a diverse range of investors.

- “Equity tranche”: The equity tranche is the tranche that absorbs the first loss (and thus is the most risky tranche) is often called an equity tranche. The remaining tranches are called mezzanine or senior tranches.

- “Junior tranche”: A junior tranche is less risky than an equity tranche but more than a senior tranche.

- “Mezzanine tranche”: Another calling for junior tranche.

- “Senior tranche”: A senior tranche is the highest tranche of a security, i.e. the one deemed least risky. Any losses on the value of the security are only experienced in the senior tranche once all other tranches have lost all their value. For this safety, the senior tranche pays the lowest rate of interest.

- “ Repurchase agreement ” or “ Repo ”: A repurchase agreement (repo) is a form of short-term borrowing for dealers in government securities. In the case of a repo, a dealer sells government securities to investors, usually on an overnight basis, and buys them back the following day.

C2 - Dictionary of traditional banking words

- “Deposit”: money placed into banking institutions for safekeeping.

- “Savings account”: an interest-bearing deposit account held at a bank or other financial institution.

- “Savings certificate”: receipt issued by a savings institution (bank, building society, credit union, etc.) to certify the ownership of a fixed or time deposit.

- “Chequing account or Checking account”: deposit account held at a financial institution that allows withdrawals and deposits.

- “Certificate of deposit”: a product offered by banks and credit unions that provides an interest rate premium in exchange for the customer agreeing to leave a lump-sum deposit untouched for a predetermined period of time.

- “Deposit insurance”: an insurance which guarantees the customer of a bank a certain amount if a bank run occurs.

- “Discount window”: an instrument of monetary policy (usually controlled by central banks) that allows eligible institutions to borrow money from the central bank, usually on a short-term basis, to meet temporary shortages of liquidity caused by internal or external disruptions.

- “Reserve requirement”: amount of funds that a bank holds in reserve to ensure that it is able to meet liabilities in case of sudden withdrawals.

- “Tier 1”: Tier 1 capital is used to describe the capital adequacy of a bank and refers to core capital that includes equity capital and disclosed reserves. Tier 1 capital is essentially the most perfect form of a bank’s capitalthe money the bank has stored to keep it functioning through all the risky transactions it performs, such as trading/investing and lending.

- “Tier 2”: Tier 2 capital is the secondary component of bank capital, in addition to Tier 1 capital, that makes up a bank’s required reserves. Tier 2 capital is designated as supplementary capital and is composed of items such as revaluation reserves, undisclosed reserves, hybrid instruments, and subordinated term debt.

- “Liquidity coverage ratio”: It refers to the proportion of highly liquid assets held by financial institutions, to ensure their ongoing ability to meet short-term obligations.

- “Net stable funding ratio”: as the liquidity coverage ratio,the net stable funding ratio is a liquidity standard. It aims to promote resilience over a longer time horizon by creating incentives for banks to fund their activities with more stable sources of funding on an ongoing basis.

- “Basel Committee on Banking Supervision”: it is the primary global standard setter for the prudential regulation of banks and provides a forum for regular cooperation on banking supervisory matters.

- "High-quality liquid assets": High-quality liquid assets are comprised of Level 1 and Level 2 assets. Level 1 assets generally include cash, central bank reserves, and certain marketable securities backed by sovereigns and central banks. Level 2 assets include, for example, certain government securities, covered bonds and corporate debt securities.

- "Government debt": debt issued by the national government in order to finance the issuing country's growth and development.

- "Corporate debt": type of debt security that is issued by a firm and sold to investors.

- "Comprehensive Capital Analysis and Review": a stress-test regime for large US banks. It aims to establish whether lenders have enough capital to cope with a severe economic shock, and assesses their risk modelling practices.

Bibliographie

Bibliography

Tobias Adrian and Adam B Ashcraft. Shadow banking regulation. *Annu. Rev. Financ. Econ.*, 4(1):99–140, 2012.

Tobias Adrian and Adam B Ashcraft. Shadow banking: a review of the literature. In *Banking crises*, pages 282–315. Springer, 2016.

Tobias Adrian and Hyun Song Shin. The shadow banking system: implications for financial regulation. *FRB of New York Staff Report*, (382), 2009.

Andrew Ang, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang. The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1):259–299, 2006a.

Andrew Ang, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang. The Cross-Section of Volatility and Expected Returns. *The Journal of Finance*, 61(1):259–299, 2006b. ISSN 1540-6261. doi: 10.1111/j.1540-6261.2006.00836.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2006.00836.x.

Werner Antweiler and Murray Z Frank. Is all that talk just noise? the information content of internet stock message boards. *The Journal of finance*, 59(3):1259–1294, 2004.

George O. Aragon and Philip E. Strahan. Hedge funds as liquidity providers: Evidence from the Lehman bankruptcy. *Journal of Financial Economics*, 103(3):570–587, March 2012. ISSN 0304-405X. doi: 10.1016/j.jfineco.2011.10.004. URL http://www.sciencedirect.com/science/article/pii/S0304405X11002364.

William C Barbee Jr, Sandip Mukherji, and Gary A Raines. Do sales–price and debt–equity explain stock returns better than book–market and firm size? *Financial Analysts Journal*, 52(2):56–60, 1996.

Nicholas C Barberis and Richard H Thaler. A survey of behavioral finance," in constantinides, george, harris, milton and rene stulz (eds.), handbook of the economics of finance, amsterdam: North-holland. 2003.

Sanjoy Basu. Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *The journal of Finance*, 32(3):663–682, 1977.

Taylor A Begley and Amiyatosh Purnanandam. Design of financial securities: Empirical evidence from private-label rmbs deals. *The Review of Financial Studies*, 30(1):120–161, 2017.

Stephen A. Berkowitz, Dennis E. Logue, and Eugene A. Noser. The Total Cost of Transactions on the NYSE. *The Journal of Finance*, 43(1):97–112, 1988. ISSN 1540-6261. doi: 10.1111/j.1540-6261.1988.tb02591.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1988.tb02591.x.

Laxmi Chand Bhandari. Debt/equity ratio and expected common stock returns: Empirical evidence. *The journal of finance*, 43(2):507–528, 1988.

Ravi Bhushan. An informational efficiency perspective on the post-earnings announcement drift. *Journal of Accounting and Economics*, 18(1):45–65, July 1994. ISSN 0165-4101. doi: 10.1016/0165-4101(94)90018-3. URL http://www.sciencedirect.com/science/article/pii/0165410194900183.

Marshall E Blume and Frank Husic. Price, beta, and exchange listing. *The Journal of Finance*, 28(2):283–299, 1973.

Financial Stability Board. Strengthening oversight and regulation of shadow banking: policy framework for addressing shadow banking risks in securities lending and repos. *August*, 29:2013, 2013.

Tim Bollerslev and Viktor Todorov. Tails, Fears, and Risk Premia. *Journal of Finance*, 66(6):2165–2211, 2011. URL http://ideas.repec.org/a/bla/jfinan/v66y2011i6p2165-2211.html.

Tim Bollerslev, Viktor Todorov, and Lai Xu. Tail risk premia and return predictability. *Journal of Financial Economics*, 118(1):113–34, 2015.

Michael W. Brandt, Pedro Santa-Clara, and Rossen Valkanov. Parametric Portfolio Policies: Exploiting Characteristics in the Cross-Section of Equity Returns. *Rev Financ Stud*, 22 (9):3411–3447, September 2009. ISSN 0893-9454. doi: 10.1093/rfs/hhp003. URL https://academic.oup.com/rfs/article/22/9/3411/1572695.

Markus K. Brunnermeier and Lasse Heje Pedersen. Market Liquidity and Funding Liquidity. *Rev Financ Stud*, 22(6):2201–2238, June 2009. ISSN 0893-9454. doi: 10.1093/rfs/hhn098. URL https://academic.oup.com/rfs/article/22/6/2201/1592184.

Greg Buchak, Gregor Matvos, Tomasz Piskorski, and Amit Seru. Fintech, regulatory arbitrage, and the rise of shadow banks. *Journal of Financial Economics*, 130(3):453–483, 2018.

Louis KC Chan, Yasushi Hamao, and Josef Lakonishok. Fundamentals and stock returns in japan. *The journal of finance*, 46(5):1739–1764, 1991.

Long Chen, Robert Novy-Marx, and Lu Zhang. An Alternative Three-Factor Model. SSRN Scholarly Paper ID 1418117, Social Science Research Network, Rochester, NY, April 2011. URL https://papers.ssrn.com/abstract=1418117.

Yongok Choi, Stefan Jacewitz, and Joon Y Park. A reexamination of stock return predictability. *Journal of Econometrics*, 192(1):168–189, 2016.

Michael J. Cooper, Huseyin Gulen, and Michael J. Schill. Asset Growth and the Cross-Section of Stock Returns. *The Journal of Finance*, 63(4):1609–1651, 2008. ISSN 1540-6261. doi: 10.1111/j.1540-6261.2008.01370.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2008.01370.x.

Shane A Corwin and Paul Schultz. A simple way to estimate bid-ask spreads from daily high and low prices. *Journal of Finance*, 2011.

Zhi Da, Qianqiu Liu, and Ernst Schaumburg. A Closer Look at the Short-Term Return Reversal. *Management Science*, 60(3):658–674, September 2013. ISSN 0025-1909. doi: 10.1287/mnsc.2013.1766. URL https://pubsonline.informs.org/doi/abs/10.1287/mnsc.2013.1766.

Kent Daniel and Sheridan Titman. Market reactions to tangible and intangible information. *The Journal of Finance*, 61(4):1605–1643, 2006.

Sanjiv R Das and Mike Y Chen. Yahoo! for amazon: Sentiment extraction from small talk on the web. *Management science*, 53(9):1375–1388, 2007.

Vinay T Datar, Narayan Y Naik, and Robert Radcliffe. Liquidity and stock returns: An alternative test. *Journal of Financial Markets*, 1(2):203–219, 1998.

Werner FM De Bondt and Richard Thaler. Does the stock market overreact? *The Journal of finance*, 40(3):793–805, 1985.

Yongheng Deng, John M Quigley, and Robert Van Order. Mortgage terminations, heterogeneity and the exercise of mortgage options. *Econometrica*, 68(2):275–307, 2000.

Joseph Engelberg. Costly information processing: Evidence from earnings announcements. In *AFA 2009 San Francisco meetings paper*, 2008.

Joseph Engelberg, R David McLean, and Jeffrey Pontiff. Anomalies and news. *The Journal of Finance*, 73(5):1971–2001, 2018.

Patricia M Fairfield, J Scott Whisenant, and Teri Lombardi Yohn. Accrued earnings and growth: Implications for future profitability and market mispricing. *The accounting review*, 78(1):353–371, 2003.

Eugene F Fama. Efficient capital markets: Ii. *The journal of finance*, 46(5):1575–1617, 1991.

Eugene F Fama. Market efficiency, long-term returns, and behavioral finance. *Journal of financial economics*, 49(3):283–306, 1998.

Eugene F Fama and Kenneth R French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56, 1993.

Eugene F Fama and James D MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3):607–636, 1973.

Jean-Sbastien Fontaine and René Garcia. Bond Liquidity Premia. *Rev Financ Stud*, 25 (4):1207–1254, April 2012. ISSN 0893-9454. doi: 10.1093/rfs/hhr132. URL https://academic.oup.com/rfs/article/25/4/1207/1579996.

George Foster, Chris Olsen, and Terry Shevlin. Earnings releases, anomalies, and the behavior of security returns. *Accounting Review*, pages 574–603, 1984.

Andrea Frazzini and Lasse Heje Pedersen. Betting against beta. *Journal of Financial Economics*, 111(1):1–25, January 2014. ISSN 0304-405X. doi: 10.1016/j.jfineco.2013.10.005. URL http://www.sciencedirect.com/science/article/pii/S0304405X13002675.

Diego Garcia. Sentiment during recessions. *The Journal of Finance*, 68(3):1267–1300, 2013.

Nicolae Garleanu and Lasse Heje Pedersen. Margin-based Asset Pricing and Deviations from the Law of One Price. *Rev Financ Stud*, 24(6):1980–2022, June 2011. ISSN 0893-9454. doi: 10.1093/rfs/hhr027. URL https://academic.oup.com/rfs/article/24/6/1980/1588773.

John Geanakoplos. The leverage cycle. NBER Macroeconomics Annual 2009, volume 24, University of Chicago Press, Chicago, IL, 2010.

A. E. Gelfand and D. K. Dey. Bayesian Model Choice: Asymptotics and Exact Calculations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56(3):501–514, 1994. ISSN 0035-9246. URL https://www.jstor.org/stable/2346123.

Lawrence R. Glosten and Paul R. Milgrom. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics*, 14(1): 71–100, March 1985. ISSN 0304-405X. doi: 10.1016/0304-405X(85)90044-3. URL http://www.sciencedirect.com/science/article/pii/0304405X85900443.

Benjamin Golez, Jens Carsten Jackwerth, and Anna Slavutskaya. Funding Illiquidity Implied by S&P 500 Derivatives. SSRN Scholarly Paper ID 2640697, Social Science Research Network, Rochester, NY, February 2018. URL https://papers.ssrn.com/abstract=2640697.

Gary Gorton and Andrew Metrick. Securitized banking and the run on repo. *Journal of Financial economics*, 104(3):425–451, 2012.

Gary Gorton, Andrew Metrick, Andrei Shleifer, and Daniel K Tarullo. Regulating the shadow banking system [with comments and discussion]. *Brookings papers on economic activity*, pages 261–312, 2010.

Ruslan Y. Goyenko, Craig W. Holden, and Charles A. Trzcinka. Do liquidity measures measure liquidity? *Journal of Financial Economics*, 92(2):153–181, May 2009. ISSN 0304-405X. doi: 10.1016/j.jfineco.2008.06.002. URL http://www.sciencedirect.com/science/article/pii/S0304405X09000099.

D. Gromb and D. Vayanos. A model of financial market liquidity based on intermediary capital. *Journal of the European Economic Association*, 8(2-3):456–466, 2010.

Denis Gromb and Dimitri Vayanos. Equilibrium and welfare in markets with financially constrained arbitrageurs. *Journal of Financial Economics*, 66(2):361–407, November 2002. ISSN 0304-405X. doi: 10.1016/S0304-405X(02)00228-3. URL http://www.sciencedirect.com/science/article/pii/S0304405X02002283.

Sanford J. Grossman and Merton H. Miller. Liquidity and Market Structure. *The Journal of Finance*, 43(3):617–633, 1988. ISSN 1540-6261. doi: 10.1111/j.1540-6261.1988.tb04594.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1988.tb04594.x.

Kathleen Weiss Hanley and Gerard Hoberg. Dynamic interpretation of emerging risks in the financial sector. *The Review of Financial Studies*, 32(12):4543–4603, 2019.

Lawrence Harris. Statistical Properties of the Roll Serial Covariance Bid/Ask Spread Estimator. *The Journal of Finance*, 45(2):579–590, 1990. ISSN 1540-6261. doi: 10.1111/j.1540-6261.1990.tb03704.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1990.tb03704.x.

Joel Hasbrouck. Liquidity in the Futures Pits: Inferring Market Dynamics from Incomplete Data. *The Journal of Financial and Quantitative Analysis*, 39(2):305–326, 2004. ISSN 0022-1090. URL https://www.jstor.org/stable/30031857.

Joel Hasbrouck. Trading Costs and Returns for U.S. Equities: Estimating Effective Costs from Daily Data. *The Journal of Finance*, 64(3):1445–1477, 2009. ISSN 1540-6261. doi: 10.1111/j.1540-6261.2009.01469.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2009.01469.x.

Patric H Hendershott and William R Schultz. Equity and nonequity determinants of fha single-family mortgage foreclosures in the 1980s. *Real Estate Economics*, 21(4):405–430, 1993.

Elaine Henry. Are investors influenced by how earnings press releases are written? *The Journal of Business Communication (1973)*, 45(4):363–407, 2008.

Steven L. Heston and Ronnie Sadka. Seasonality in the cross-section of stock returns. *Journal of Financial Economics*, 87(2):418–445, February 2008. ISSN 0304-405X. doi:

10.1016/j.jfineco.2007.02.003. URL http://www.sciencedirect.com/science/article/pii/S0304405X0700195X.

Steven L Heston and Nitish Ranjan Sinha. News vs. sentiment: Predicting stock returns from news stories. *Financial Analysts Journal*, 73(3):67–83, 2017.

Alexander Hillert, Heiko Jacobs, and Sebastian Müller. Media makes momentum. *The Review of Financial Studies*, 27(12):3467–3501, 2014.

David Hirshleifer, Kewei Hou, Siew Hong Teoh, and Yinglei Zhang. Do investors overvalue firms with bloated balance sheets? *Journal of Accounting and Economics*, 38:297–331, December 2004. ISSN 0165-4101. doi: 10.1016/j.jacceco.2004.10.002. URL http://www.sciencedirect.com/science/article/pii/S0165410104000795.

G. X. Hu, J. Pan, and J. Wang. Noise as information for illiquidity. *The Journal of Finance*, 68(6):2341–2382, 2013. ISSN 1540-6261. doi: 10.1111/jofi.12083. URL http://dx.doi.org/10.1111/jofi.12083.

Roger D. Huang and Hans R. Stoll. Market Microstructure and Stock Return Predictions. *Rev Financ Stud*, 7(1):179–213, January 1994. ISSN 0893-9454. doi: 10.1093/rfs/7.1.179. URL https://academic.oup.com/rfs/article/7/1/179/1568543.

Heiko Jacobs and Sebastian Müller. Anomalies across the globe: Once public, no longer existent? *Journal of Financial Economics*, 135(1):213–230, 2020.

Harold Jeffreys. *The Theory of Probability*. OUP Oxford, August 1998. ISBN 978-0-19-158967-6. Google-Books-ID: vh9Act9rtzQC.

Narasimhan Jegadeesh. Evidence of predictable behavior of security returns. *The Journal of finance*, 45(3):881–898, 1990.

Narasimhan Jegadeesh and Sheridan Titman. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, 48(1):65–91, 1993a.

Narasimhan Jegadeesh and Sheridan Titman. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance*, 48(1):65–91, 1993b. ISSN 1540-6261. doi: 10.1111/j.1540-6261.1993.tb04702.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1993.tb04702.x.

Wei Jiang, Ashlyn Aiko Nelson, and Edward Vytlacil. Securitization and loan performance: Ex ante and ex post relations in the mortgage market. *The Review of Financial Studies*, 27(2):454–483, 2014.

Jonathan M. Karpoff and Ralph A. Walkling. Short-term trading around ex-dividend days: Additional evidence. *Journal of Financial Economics*, 21(2):291–298, September 1988. ISSN 0304-405X. doi: 10.1016/0304-405X(88)90063-3. URL http://www.sciencedirect.com/science/article/pii/0304405X88900633.

Dongcheol Kim and Myungsun Kim. A multifactor explanation of post-earnings announcement drift. *Journal of Financial and Quantitative Analysis*, 38(2):383–398, 2003.

Pter Kondor and Dimitri Vayanos. Liquidity Risk and the Dynamics of Arbitrage Capital. *The Journal of Finance*, 74(3):1139–1173, 2019. ISSN 1540-6261. doi: 10.1111/jofi.12757. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12757.

Serhiy Kozak, Stefan Nagel, and Shrihari Santosh. Shrinking the cross-section. *Journal of Financial Economics*, 2019.

Albert S. Kyle. Continuous Auctions and Insider Trading. *Econometrica*, 53(6):1315–1335, 1985. ISSN 0012-9682. doi: 10.2307/1913210. URL https://www.jstor.org/stable/1913210.

Josef Lakonishok, Andrei Shleifer, and Robert W Vishny. Contrarian investment, extrapolation, and risk. *The journal of finance*, 49(5):1541–1578, 1994.

Mark Lang and Russell Lundholm. Cross-Sectional Determinants of Analyst Ratings of Corporate Disclosures. *Journal of Accounting Research*, 31(2):246–271, 1993. ISSN 0021-8456. doi: 10.2307/2491273. URL https://www.jstor.org/stable/2491273.

David A. Lesmond, Joseph P. Ogden, and Charles A. Trzcinka. A New Estimate of Transaction Costs. *Rev Financ Stud*, 12(5):1113–1141, October 1999. ISSN 0893-9454. doi: 10.1093/rfs/12.5.1113. URL https://academic.oup.com/rfs/article/12/5/1113/1565408.

David A. Lesmond, Michael J. Schill, and Chunsheng Zhou. The illusory nature of momentum profits. *Journal of Financial Economics*, 71(2):349–380, February 2004. ISSN 0304-405X. doi: 10.1016/S0304-405X(03)00206-X. URL http://www.sciencedirect.com/science/article/pii/S0304405X0300206X.

Feng Li. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and economics*, 45(2-3):221–247, 2008.

John Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, 47(1):13–67, 1965.

Jianan Liu, Robert F Stambaugh, and Yu Yuan. Absolving beta of volatilitys effects. *Journal of Financial Economics*, 128(1):1–15, 2018.

Igor Livshits, James MacGee, and Michele Tertilt. Consumer bankruptcy: A fresh start. *American Economic Review*, 97(1):402–418, 2007.

Tim Loughran and Bill McDonald. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65, 2011.

David Luttrell, Harvey Rosenblum, Jackson Thies, et al. Understanding the risks inherent in shadow banking: A primer and practical lessons learned. *Staff papers*, (Nov), 2012.

R David McLean and Jeffrey Pontiff. Does academic research destroy stock return predictability? *The Journal of Finance*, 71(1):5–32, 2016.

Tobias J Moskowitz and Mark Grinblatt. Do industries explain momentum? *The Journal of finance*, 54(4):1249–1290, 1999a.

Tobias J. Moskowitz and Mark Grinblatt. Do Industries Explain Momentum? *The Journal of Finance*, 54(4):1249–1290, 1999b. ISSN 1540-6261. doi: 10.1111/0022-1082.00146. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/0022-1082.00146.

Stefan Nagel. Evaporating Liquidity. *Rev Financ Stud*, 25(7):2005–2039, July 2012. ISSN 0893-9454. doi: 10.1093/rfs/hhs066. URL https://academic.oup.com/rfs/article/25/7/2005/1602153.

Andy Naranjo, M Nimalendran, and Mike Ryngaert. Stock returns, dividend yields, and taxes. *The Journal of Finance*, 53(6):2029–2057, 1998.

Benjamin Nelson, Gabor Pinter, and Konstantinos Theodoridis. Do contractionary monetary policy shocks expand shadow banking? *Journal of Applied Econometrics*, 33(2):198–211, 2018.

Jeffrey Ng, Tjomme O. Rusticus, and Rodrigo S. Verdi. Implications of Transaction Costs for the PostEarnings Announcement Drift. *Journal of Accounting Research*, 46(3):661–696, 2008. ISSN 1475-679X. doi: 10.1111/j.1475-679X.2008.00290.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-679X.2008.00290.x.

Robert Novy-Marx. The other side of value: The gross profitability premium. *Journal of Financial Economics*, 108(1):1–28, April 2013. ISSN 0304-405X. doi: 10.1016/j.jfineco.2013.01.003. URL http://www.sciencedirect.com/science/article/pii/S0304405X13000044.

Robert Novy-Marx and Mihail Velikov. A Taxonomy of Anomalies and Their Trading Costs. *Rev Financ Stud*, 29(1):104–147, January 2016. ISSN 0893-9454. doi: 10.1093/rfs/hhv063. URL https://academic.oup.com/rfs/article/29/1/104/1844518.

Yang-Ho Park. Price dislocation and price discovery in the s&p 500 options and vix derivatives markets, 2015.

Andrew J. Patton and Brian Weller. What You See Is Not What You Get: The Costs of Trading Market Anomalies. SSRN Scholarly Paper ID 3034796, Social Science Research Network, Rochester, NY, May 2019. URL https://papers.ssrn.com/abstract=3034796.

Joseph D. Piotroski. Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers. SSRN Scholarly Paper ID 249455, Social Science Research Network, Rochester, NY, January 2001. URL https://papers.ssrn.com/abstract=249455.

Jeffrey Pontiff and Artemiza Woodgate. Share issuance and cross-sectional returns. *The Journal of Finance*, 63(2):921–945, 2008.

Zoltan Pozsar. The rise and fall of the shadow banking system. *Regional Financial Review*, 44:1–13, 2008.

Zoltan Pozsar, Tobias Adrian, Adam Ashcraft, and Hayley Boesky. Shadow banking. *New York*, 458(458):3–9, 2010.

Richard Roll. A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market. *The Journal of Finance*, 39(4):1127–1139, 1984. ISSN 1540-6261. doi: 10.1111/j.1540-6261.1984.tb03897.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1984.tb03897.x.

William F Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442, 1964.

Andrei Shleifer and Robert W. Vishny. The Limits of Arbitrage. *The Journal of Finance*, 52 (1):35–55, 1997. ISSN 1540-6261. doi: 10.1111/j.1540-6261.1997.tb03807.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1997.tb03807.x.

Richard G Sloan. Do stock prices fully reflect information in accruals and cash flows about future earnings? *Accounting review*, pages 289–315, 1996.

Mark T Soliman. The use of dupont analysis by market participants. *The Accounting Review*, 83(3):823–853, 2008.

Paul C Tetlock. Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, 62(3):1139–1168, 2007.

Paul C Tetlock, Maytal Saar-Tsechansky, and Sofus Macskassy. More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, 63(3):1437–1467, 2008.

Brian M. Weller. Measuring Tail Risks at High Frequency. *Rev Financ Stud*, 32(9):3571–3616, September 2019. ISSN 0893-9454. doi: 10.1093/rfs/hhy133. URL https://academic.oup.com/rfs/article/32/9/3571/5244643.

Robert E. Whaley. The Investor Fear Gauge. *The Journal of Portfolio Management*, 26 (3):12–17, April 2000. ISSN 0095-4918, 2168-8656. doi: 10.3905/jpm.2000.319728. URL https://jpm.pm-research.com/content/26/3/12.

Kairong Xiao. Monetary transmission through shadow banks. *The Review of Financial Studies*, 33(6):2379–2420, 2020.

Yuhang Xing. Interpreting the value effect through the q-theory: An empirical investigation. *The Review of Financial Studies*, 21(4):1767–1795, 2007.

Min Zhu. Jackknife for bias reduction in predictive regressions. *Journal of Financial Econometrics*, 11(1):193–220, 2012.